

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model.

Run on: April 29, 2003, 17:10:24 ; Search time 11.7002 Seconds
(without alignments)
3367.676 Million cell updates/sec

Title: US-10-009-332-1
Perfect score: 5164
Sequence: 1 MLLGIITLAFAGTAGTGF.....DQCNLHRKPQELDFCVLRPC 950

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	2482.5	48.1	968	1	ATSL_MOUSE		P97857 mus musculus
2	2480.5	48.0	967	1	ATSL_HUMAN		Q9uh18 homo sapien
3	2471.1	47.9	967	1	ATSL_RAT		Q9wug1 rattus norv
4	2135.5	41.4	890	1	ATSL_HUMAN		Q9up79 homo sapien
5	2113.4	40.9	905	1	ATSL_MOUSE		P57110 mus musculus
6	1965.1	38.1	1629	1	ATSL_HUMAN		Q9p2n4 homo sapien
7	1917.3	37.1	837	1	ATSL_HUMAN		Q75173 homo sapien
8	1856.3	35.9	930	1	ATSL_MOUSE		Q9r001 mus musculus
9	1847.5	35.8	930	1	ATSL_HUMAN		Q9una0 homo sapien
10	1745.5	33.8	630	1	ATSL_RAT		Q9esp7 rattus norv
11	1396.2	25.1	1077	1	ATSL_HUMAN		Q9h324 homo sapien
12	1275.2	24.7	1593	1	ATSL_HUMAN		P58397 homo sapien
13	1230.2	23.8	1211	1	ATSL_HUMAN		P78325 mus musculus
14	1212.5	23.5	997	1	ATSL_HUMAN		Q95450 h adamts-2
15	1192.2	23.1	1205	1	ATSL_BOVIN		Q9ukp4 homo sapien
16	1183.2	22.9	1205	1	ATSL_HUMAN		P79331 b adamts-2
17	1048.2	20.3	860	1	ATSL_HUMAN		O15072 homo sapien
18	601.5	11.6	245	1	ATSL_BOVIN		Q9ukp5 homo sapien
19	599.1	11.6	207	1	ATSL_BOVIN		Q9tt93 bos taurus
20	381.5	7.4	450	1	ATSL_MOUSE		Q9tt92 bos taurus
21	380.5	7.4	824	1	AD08_HUMAN		P58459 mus musculus
22	362.7	7.0	776	1	AD28_MACFA		P78325 mus musculus
23	362.7	7.0	956	1	AD19_HUMAN		Q9xsl6 macaca fasc
24	349.5	6.8	813	1	AD33_HUMAN		Q9h013 homo sapien
25	344.5	6.7	571	1	DLST_BOVTA		Q9bz11 homo sapien
26	342.5	6.6	774	1	AD28_MOUSE		P30431 bothrops ja
27	340.6	6.6	775	1	AD28_HUMAN		Q9j1n6 mus musculus
28	337.6	6.5	857	1	AD22_MOUSE		Q9ukq2 homo sapien
29	332.6	6.4	906	1	AD22_HUMAN		Q9riv6 mus musculus
30	329.5	6.4	920	1	AD19_MOUSE		Q9p0k1 homo sapien
31	329.5	6.4	903	1	AD12_MOUSE		Q35674 mus musculus
32	321.6	6.2	826	1	AD08_MOUSE		Q61824 mus musculus
33	314.6	6.1	909	1	AD12_HUMAN		Q05910 mus musculus
							Q43184 homo sapien

34 308 6.0 789 1 AD07_RAT Q63180 rattus norv
35 297 5.8 480 1 DISA_TRIGA P15503 trimeresuru
36 296 5.7 413 1 ACLA_AKAC Q9pw35 agkistrodon
37 288.5 5.6 788 1 AD07_MOUSE O35227 mus musculus
38 286 5.5 769 1 AD11_HUMAN O75078 mus sapien
39 284 5.5 414 1 HRTD_CROAT P15167 crotalus at
40 277.5 5.4 814 1 AD15_HUMAN Q13444 homo sapien
41 277 5.4 773 1 AD11_MOUSE O42596 xenopus lae
42 274.5 5.3 935 1 AD22_XENLA Q94182 crotalus lae
43 273.5 5.3 478 1 HRTD_CROAT P34182 crotalus at
44 268 5.2 797 1 AD33_MOUSE Q923w9 mus musculus
45 260.5 5.0 754 1 AD07_HUMAN Q9h2u9 homo sapien

ALIGNMENTS

RESULT 1
ATSL_MOUSE
ID ATSL_MOUSE STANDARD; PRT; 968 AA.
AC P97857; O54768;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN ADAMTS1
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RX MEDLINE=98110583; PubMed=9441751;
RA Kuno K., Lizasa H., Ohno S., Matsushima K.;
RT "The exon/intron organization and chromosomal mapping of the mouse
RT ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";
RL Genomics 46:466-471(1997).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97150761; PubMed=8995297;
RA Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.,
RA Matsushima K.;
RT "Molecular cloning of a gene encoding a new type of metalloproteinase-
RT disintegrin family protein with thrombospondin motifs as an
RT inflammation associated gene.";
RL J. Biol. Chem. 272:556-562(1997).
[3]
RP CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.
RX MEDLINE=99303657; PubMed=10373500;
RA Kuno K., Terashima Y., Matsushima K.;
RT "ADAMTS-1 is an active metalloproteinase associated with the
RT extracellular matrix.";
RL J. Biol. Chem. 274:18821-18826(1999).
[4]
RP FUNCTION.
RX MEDLINE=20389568; PubMed=10930576;
RA Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.,
RA Ohno H., Matsushima K.;
RT "ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";
RL FEBS Lett. 478:241-245(2000).
[5]
RP FUNCTION, AND INDUCTION
RX MEDLINE=20243757; PubMed=10781075;
RA Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W.,
RA Richards J.S.;
RT "progesterone-regulated genes in the ovulation process: ADAMTS-1 and
RT cathepsin L proteases.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).
CC -1- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH

CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).

CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1691-GLU-I-LEU-1692
CC SITE. WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.

CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX.

CC -1- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY
CC INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY
CC LIPOPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA
CC CELLS OF PREOVULATORY FOLLICLES.

CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FORIN ENDOPEPTIDASE.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.

CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 7.

CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC -----
CC EMBL; AB001735; BAA24501.1; ALT-INIT.
CC EMBL; D67076; BAA11088.1; ALT_FRAME.
CC MEROPS; M12.222; .
CC MGD; MGI:109249; Adamts1.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; Zn_MTpeptdse.
CC Pfam; PF00090; Tsp_1; 3.
CC Pfam; PF01421; Reprolysin; 1.
CC Pfam; PF01562; Rep_M12B_propep; 1.
CC SMART; SM00209; TSP1; 3.
CC PROSITE; PS0215; ADAM_MEPPO; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE; PS00092; TSP1; 3.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
KW SIGNAL; 1 48
KW SIGNAL; 1 48
FT PROPEP 49 253
FT CHAIN 254 968
FT SITE 206 206
FT METAL 402 402
FT METAL 403 403
FT ACT_SITE 406 406
FT METAL 412 412
FT METAL 477 559
FT DOMAIN 560 617
FT DOMAIN 618 725
FT DOMAIN 726 850
FT DOMAIN 851 909
FT DOMAIN 910 968
FT DOMAIN 195 199
FT CARBOHYD 548 548
FT CARBOHYD 721 721
FT CARBOHYD 765 765
FT CARBOHYD 783 783
FT CARBOHYD 946 946
FT MUTAGEN 403 403
FT CONFLICT 335 335
FT CONFLICT 425 425
SQ SEQUENCE 968 AA; 105841 MW; 42EBDA55499FB6C1 CRC64;

Query Match	48.1%	Score 2482.5;	DB 1;	Length 968;
Best Local Similarity	48.0%	Pred. No. 1.3e-167;		
Matches 478;	Conservative 158;	Mismatches 249;	Indels 111;	Gaps 22;

QY	1	M L L G I L Y L A F A G R T A G G - F E P E R E V V P I R L P D I N G R R Y Y W R G P - E D S G D O G L I F O I	57
DB	37	L I L L A S I T M L L C A R G A H G R P T E E D E L V L P - S L E - - - - - R A P G H D S T T T R L - - R L	83
QY	58	T A F Q D E F Y L H L T P D A O F L A P A F S T E H L G V P L Q G L T G G S - - - - - S D L R C F Y S G D V A	109
DB	84	D A F G Q L L H L K I Q P D S G F L A G E T L Q V - - - - - G R S P G S E A Q H L D P T G L A H C F Y S G T V N G	138
QY	110	E P D S F A A V S L C G L R G A F G Y G A E V I S P L P N A S - - - - - A P A A Q R N S Q A - - - - - H L Q R R G	161
DB	139	D P G S A A A L S L C E G V R G A F Y L Q G E E F I Q P A G V A T E R L A P A V E E S A R P Q F H I L R R R -	197
QY	162	V P G G S G D P T S C G V A S G W N P A I L R A L D P Y K P R R A G F G E S R R R - - - - - - - - - - - - -	206
DB	198	- - - - - R G S G G A K G C V M D -	243
QY	207	- - - - - S G R A K R F V S I P R Y V E T L V V A D E S M Y K F H G A D L E H Y L L T L A T A A R L Y R H P S I L N P I	262
DB	244	S G P G S I R K R F V S P R Y V E T L V A D Q S M A D F H G S G L K H Y L L T L F S V A A R F Y K H P S I R N S I	303
QY	263	N I V V K V L L R D R D S G P K V T G N A A L T L R N F C A W K K L N K V S D K H P E Y W D T A I L F T R Q D L C	322
DB	304	S L V V K I L Y I E Q K G P E V T S N A A L T L R N F C W Q K H N S P S D R D P E H Y D T A I L F T R Q D L C	363
QY	323	G A T T C D T L G M A D V G T M C D P K R S C V I E D D G L P S A F T T A H E L G H V F M P H D N V K V C E E V F G	382
DB	364	G S H F C D T L G M A D V G T V C D P S R S C S V I E D D G L Q A A F T T A H E L G H V F M P H D D A K H C A S L N G	423
QY	383	K L R A N H M S P T L I Q I D R A N P W S A C S A A I T D F L D S H G D C L L D Q S K P L S L E D L P G A S Y	442
DB	424	V T G D S H L M A S M L S L D H S Q P W S P C S A Y M V T S F L D N G H G E C L M D K P N P I K L S D L P G T I L Y	483
QY	443	T L S Q O C E L A F G V G S K P C P - Y M Q Y C T K L W C T G R A K A G M V C Q T R H F P W A D T S C G E G K L C L K	501
DB	484	D A N R Q C O F T F G E S K H C P D A A S T C T T L W C T G T S G G L L V C Q T K H F P W A D T S C G E G K W C V S	543
QY	502	G A C V E R H N L N K H - - - - - R V D G S W A K W D P G P C S R T C G G V Q L A R Q C T N P T P A N G K Y C E G V	558
DB	544	K V C N K T D M - K H E A T P V H G S W G P W G D C S R T C G G V Q Y T M R E C D N P V P K N G K Y C E G K	602
QY	559	R V K Y R S C N L E P C P S S A S G S K F R E E Q C E A F N G Y N H S N R L T L A V A W P K Y S G V S P R D K C K L	618
DB	603	R V R Y R S C N I E D C P D N - N G K T F R E E Q C E A H N E F S K A S F G N E P T V E W T P K Y A G Y S P D R C K L	661
QY	619	I C R A N G T G Y F Y L A P K V D G T L C S P D S T S Y C V Q G K I K A C D G N L G S K R F D K C G V C G G D	678
DB	662	T C E A K G I G Y F V L Q P K V D G T P C S P D S T S Y C V Q G C V K A G C D R I D S K K K F D K C G V C G G N	721
QY	679	N K S C K V T G L T K P M H G Y N F V A I P A G A S I D I R Q R G Y K G L I G D D N Y L A K N S Q G K Y L N	738
DB	722	G S T C K K M G I V T S T R P G Y H D I V T P A G A T N I E V K H R N Q R G S R N N G S F L A I R A A D G Y I I L N	781
QY	739	G H F V V S A V E R D L V V K G S L L R Y S G T G A V E S I Q A S R P L E P L T V E L S V G K M T P P R V R Y S F	798
DB	782	G N F T L S L E Q D L T Y K G T V L R Y S G S S A A L E I R S F S P L K E P L T I Q V L M V G H A L R P K I K F T Y	841
QY	799	Y L P K E P R E D K S H P K D P R G P S V L H N S Y L S L S N V Q E P D D R P P A R W A G S N G P C S A C G S G	858
DB	842	F M K K K T E S - - - - - F N A I P T F S -	874
QY	859	L Q R A V D C R S A G O R T V P A C D A A H - - - - - R P V E T Q A C G E - P C P T W E L S A W S P C S K C G R G F O R	915
DB	875	W Q R V V Q C R D I N G H - - - - - P A S E C A K E V K P A S T R P C A D I U P C P H W Q V G D W S P C S K C G K G I K K	931
QY	916	R S L K V C G G G R L L A R D C N L H R K P O E - L D F C V L R P C	950
DB	932	R T L K C V S H D G V L S N E S C D P L K K P K H Y I D C T L T Q C	967

Qy	1	MLLLGILTLAFAGRTAGGPEPEREVVVVIRLPDINGRRYRWRGPDSDGQGLIFQITAF	60
Ds	36	LLLLAAALLVADALGRPSDEDELVVP-ELE-----RAPH-----GHGTRLRHLAF	81
Qy	61	QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRCFYSGVDNAEPDSF	114
Ds	82	DQDLLELRDPSFLAPGFTLQNVGRKSGSEPLP-----ETDLAHCFFYSGVNGDPSSA	136
Qy	115	AAYSGLAGGLRGAGFYGAETVISPFLNAS---APAAQRNSQGA---HLLQ---RRGVPGG	165
Ds	137	AALSCEGVGAYLLGEAYFIQPLPAASERLATAAPGKPPAPLQFHLRLNRQGDVG	196
Qy	166	PSG-----DPTSRC-----GVASG---WNPAILRALDPYKPRRAGFGESESRRS	207
Ds	197	TCGVVDDEPRPTKGAETDEDETEGDEGPWS-----PQDPALQGVGP-TGTGS	247
Qy	208	GRAKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLTAARLYRHPSILNPINIVV	267
Ds	248	IRKRFVSSHRYVETMLVADQSAEFGSLGKHYLLTLFSVAARLYKHPHSIRNSVSLVV	307
Qy	268	KVILLRDRDGPVKVTGNAALTFRNFCAMQKKNVSKDHPKPEYWDTAITLFTRODLCGATTC	327
Ds	308	KILVIHQEKGPVTSNAALTFRNFCWQKHNPSPDRDAEHYDTAILFTRODLCGSGTC	367
Qy	328	DTLGMADVMTCDPKRSCSVIEDGLPSPAFTTAHELGHVFNPMHNVKVEEVEFGKLRAN	387
Ds	368	DTLGMADVMTCDPKRSCSVIEDGLQAFTTAHELGHVFNPMHNDKACASLNGYNQDS	427
Qy	388	HMSPTLIQIDRANPWSACSAITDFDSHGDCILDDPSKPSLSPEDLPASVYLSQ	447
Ds	428	HMAASMLSNLDHSGPWSPCSAYMITSLFNGHGECLMDKFNQFIQPLGDLPTGYDANRQ	487
Qy	448	CELAFGVSGPCP-IMOYCTKLMCTKAKQOMVCOTRHPFWDGTSCGEGKCLKGACVE	506
Ds	488	COFTFGEDSKHCPDAACTCTLWCTGSGVLVCQTKHFPWADGTSCGEGKWCINGKCVN	547
Qy	507	RHNLNKH---RVDGSNAKWDYPGCSCTCGGVQLARRQCTNTPANGKYCEGVKVR	563
Ds	548	KTD-RKHFDFTFPHGSGWGMGFWGDCSTCTCGGVQYTMRECDNVPVNGGKYCEGKVR	606
Qy	564	SCNLEPCSSASGSKFEEQCEAFNGYHNSTNRLTLAVAVPKYSGVSPRDKCKLICRAN	623
Ds	607	SCNLEDCPDN-NGKTFREEQCEAHNESKASFGSGPAVEMPKYAGVSPKDRCKLICQAK	665
Qy	624	GTGYFYVLAPKVDGTLCPSPDSTSVQGVQKICRAGCDNGNLSGKKRFDKCGVCGDNKSK	683
Ds	666	GIGYFYVLPKVDGTPCSPDSTSVQGVQCVKAGCDRIIDSKKFKDKCGVCGGNGSTCK	725
Qy	684	KVTGLFTKPMHYNFVAIPAGASSIDIRQGYKGLIGDDNYLAKNSQGYLLNGHFVV	743
Ds	726	KISGVTSAKPGYHDIITPTGATNIEVKQNRNORSGRNGSFLAIAKADGTYILNGDYTL	785
Qy	744	SAVERDLVWKSLLRYSGTGTAVESLQASRPILFTEVLSVGMKTPPPRVYSFYLPKE	803
Ds	786	STLEQIDMYKGVLYRSGSSNALERISFSLKPEPLTQIVLTVGNALRPKIKTYFYVKK	845
Qy	804	PREDKSHHPKDRGPSVLHNSVLSLNOVEQDDPPRPARVWAGSPCSCSGSLQKRA	863
Ds	846	-----KES-----FNAIPFES-----AWVIEWEGSCSKSCGELGWORRL	878
Qy	864	VDCRGSAGORTVPACDAAH---RPVETQACGE-PCPTWELSNWSPCSCSGRGRRLSKC	920
Ds	879	VECDIRNGQ---PASECAKEVKPASTRPCADHPCPQMLGSESCSKTCGKGYKRLSLAK	935
Qy	921	VHGGRLLARDQCNLHRRKQPQ-LDFCVLRPC	950
Ds	936	LSHGGVLSHESCDPLKKPKHFIDFCTMAEC	966
AC	99WUQ1; Q9ER11;	RESULT 3	
ID	ATSL_RAT	STANDARD;	
AC	Q9WUQ1; Q9ER11;	PRT;	967 AA.

16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
ADAMTS1.
Gnattus norvegicus (Rat).
Ossakaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE=Brain;
RA Liu X., Yu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A., Little S.P.;
RT "Induction of a disintegrin and metalloproteinase with the thrombospondin type I motif (ADAMTS).";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE OF 18-967 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=20304099; PubMed=10847486;
RA Diamantis I., Luethi M., Hoesli M., Reichen J.;
RT "Cloning of the rat ADAMTS-1 gene and its down regulation in endothelial cells in cirrhotic rats.";
RL Liver 20:165-172(2000).
CC -!- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY SIMILARITY). ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU-1-LEU-1684 SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).
CC -!- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM CIRRHOTIC LIVER.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL: AF149118; AAD34012.1; -
CC EMBL: AF304446; AAG29823.1; -
CC MEROPS: M12.222; -
CC InterPro: IPR001762; Disintegrin.
CC InterPro: IPR002870; Pep_M12B_propep.
CC InterPro: IPR001590; Reprolysin.
CC InterPro: IPR000884; TSPI.
CC InterPro: IPR000130; Zn_MTpeptidse.
CC Pfam: PF00090; tsp_1; 6;
CC Pfam: PF01421; Reprolysin; 2.
CC Pfam: PF01562; Pep_M12B_propep; 2.
CC SMART: SM00209; TSPI; 3.
CC SMART: PS00215; ADAM_MEPRO; 1.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC PROSITE: PS00092; TSPI; 2.
CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
CC Hydrolase; Metalloproteinase; Zinc; Signal; Glycoprotein; Zymogen; Repeat; Extracellular matrix; Heparin-binding.
KW

```
FT SIGNAL 1 54 POTENTIAL.
FT PROPEP 55 252 BY SIMILARITY.
FT CHAIN 253 967 ADAMTS-1.
FT SITE 203 205 CYSTEINE SWITCH (POTENTIAL).
FT METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 405 405 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 411 411 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 476 558 DISINTEGRIN-LIKE.
FT DOMAIN 559 615 TSP TYPE-1.
FT DOMAIN 616 724 CYS-RICH.
FT DOMAIN 725 857 SPACER.
FT DOMAIN 858 907 TSP TYPE-1 2.
FT DOMAIN 908 967 TSP TYPE-1 3.
FT DOMAIN 194 198 POLY-ARG.
FT CARBOHYD 547 577 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 21 31 I -> V (IN REF. 2).
FT CONFLICT 26 31 KFRSSQ -> RSRGSL (IN REF. 2).
FT CONFLICT 49 49 V -> A (IN REF. 2).
FT CONFLICT 72 72 R -> P (IN REF. 2).
FT CONFLICT 79 79 L -> TR (IN REF. 2).
FT CONFLICT 249 249 R -> G (IN REF. 2).
FT CONFLICT 262 265 TMLV -> NLLK (IN REF. 2).
FT CONFLICT 607 607 S -> F (IN REF. 2).
FT CONFLICT 936 936 L -> V (IN REF. 2).
FT CONFLICT 962 962 I -> T (IN REF. 2).
FT CONFLICT 967 967 AA; 105705 MW; F93C864F6DCDB4CF CRC64;
SQ SEQUENCE 967 AA; 105705 MW; F93C864F6DCDB4CF CRC64;

Query Match 47.9%; Score 2471; DB 1; Length 967;
Best Local Similarity 48.4%; Pred. No. 8.6e-167;
Matches 477; Conservative 161; Mismatches 256; Indels 92; Gaps 22;

QY 1 MILLGILTLAPAGRTAGG--FEPEERVVPIRLDPDINGRRYWRGPDGGOGLFIQIT 58
DB 37 LLLLSITMLLCVGRGARGHPTDEELVL-----PSLERARGH-----DSTILLRD 83
QY 59 AFOEDFYLHLTPAQFLAPAFSTHGLGVPLQGLTGGS-----SDLRRCFYSGDVNAE 110
DB 84 AFGQQLHLKLQPDGSLAPGFTLTQV-----GRSPGSEAHQLDPTGDLAHCFYGTGNGD 138
QY 111 PDSFAVSLCGGLRGAFGRGAEYVTSPLPNAS-----APAAQRNSQGA-----HLL--QRR 160
DB 139 PSSAAALSCEGVGAFYLGQEEFFIQAPAVATERLVPAPKEESIAAPPREHILRRRR 198
QY 161 GVPGFSGD-----PTSRCGVASGNWPAILLRALDPYKPRRAGFGESESRRRSGRAKRV 214
DB 199 GSGGACKGVNDEETLPTSSGREGSQNTPOWPLRNP--TPQGAG---KPTGPGSIRKKRV 254
QY 215 SIPRYVETLVADSMVKFPGADLEHYLLTLLAATARYRHPHSILNPNIVVVKVLLLRD 274
DB 255 SSPRYVETLVADQSMADFHGSLGHLHYLLTFSVAARFYKHPHSIRNSISLVVVKVLLIYE 314
QY 275 RDSGPVGTGNAALTNFCAWOKKLNKVSCKHPEVWDFTAILFTRODLGCAATCDTLGMAD 334
DB 315 EQKGPEVTGNAALTNFNCFSWQKHNPSDRDPEHYDTAILFTRODLGCSHTCDTLGMAD 374
QY 335 VGTMCDPKRSVIEDDGLPSAFTTAHGLGHVFNPHONVKNVCEEVFGKLRANHHMSTL 394
DB 375 VGTVCDFPSRCSVIEDDGLQAATTAHGLGHVFNPHDDAKCAFNGSVGSDSHLMAFSL 434
QY 395 IQIDRANPWSAGSAIITDPLDSHGDCLLDQPSKPSILPELDPGASVYTLSSOCELAFGV 454
DB 435 SSLDHSOPWSPCSAYMVTSLFNLNGHGECLMDKQPNIKPLPSLDPLTGLYDANRQCOFTGE 494
QY 455 GKPCP--YMOYCTKLWCTKAGQVMQVQTRHPFWADGTCGSGBKLCCLKAGCAVERHNLNKH 513
DB 495 ESTHCPDAASTCLWCTGSGLLVCQTKHPPWADGTCGSGBKWCVSGKCNKTKDM--KH 553
QY 514 ---RVDGSNAKWDPYGPCSRTCGGQVQLARRQCTNPTTANGKCYCEGVRYKRSCLNLEPC 570
```

```
DB 554 FATPVHGWGPMGWDGCDSCRTCGGVQYTMRECDNPVKNNGKCYCEGKRVYRSCNIDDC 613
QY 571 PSSASGSKFREQCEAFNGYHNHNTNRLTLAVAWVPKYSVSPROKCKLICRANGTGYFYV 630
DB 614 PDN--NGKTFREQCEAHNEFSKASFGNEPTVEWTPKYAGVPKORCKLTCEAKGLGYFFV 672
QY 631 LAPKVVDTGLSPDSTSVCVQKCLKAGCDNLSKKRFDKCGVCGGDNKCKKVTGLFT 690
DB 673 LQPKVVDTPCSPDSTSVCVQKCYKAGCDRIIDSKKFDKCGVCGGNGSTCKKISGTVT 732
QY 691 KPMHGYNEFWAIPAGASSIDIRQRYKGLIGDNDYALKNSQKYLNLNGHFWWSAVERDL 750
DB 733 STRPCYHDIIVIPAGATNIEVKNRPNRGRNNGSFLAIRAADGYIILNNGTTLSTLEODL 792
QY 751 VVKGLLRYSGTGTAVESLQASRPILPTVEVLSVGKMTPPRVYSYLPK--EPREDK 808
DB 793 IYKGTVLRYSGSSAALIRSFSPLEKPLTIQVLMVGHALRPKIKYTYFMKKKTEP---- 848
QY 809 SSHKPDPRGSPVHLNLSVLSLNOVEQPDPRPARKVAGSWGPCSNASCSGSLQKRAYDCRG 868
DB 849 -----FNAIPTS-----EWVIEWGECSKTCGSGWQRRVVECRD 883
QY 869 SAGQRTVPACDAAH--REVETQAGE--PCPTWELSAWPCSKSGRGFORSLKCVGHG 925
DB 884 INGH---PASECAKEVKPASTRPCADLPCPRWQVGDWSPCSKTCGKYKKTCLKLSHDG 940
QY 926 RLLARDQCNLHRKQOE--LDFOCVLRPC 950
DB 941 GVLNSCEDPLKPKHYIDFCILTQC 966

RESULT 4
ATS8_HUMAN
ID ATS8_HUMAN STANDARD; PRT; 890 AA.
AC Q9UP79; Q9NZ50;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2) (METH-8).
DE Homo sapiens (Human).
GN ADAMTS8 OR METH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=93367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S., Lombardo M., Iruela-Arispe M.L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity.";
RL J. Biol. Chem. 274:23349-23357(1999).
RN [2]
RP SEQUENCE OF 195-440 FROM N.A.
RX MEDLINE=20079168; PubMed=10610729;
RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on mouse chromosome 9 and human chromosome 11.";
RL Genomics 62:312-315(1999).
CC CC-1- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
CC CC-1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (by similarity).
CC CC-1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC CC-1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND KIDNEY.
CC CC-1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC CC-1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
```

CC CC SIMILARITY).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: AF060153; AAD48081.1; -

CC EMBL: AF175283; AAF25806.1; -

CC HSSP: P341179; 1IAG.

CC MEROPS: M12.26; -

CC MIM: 605175; -

CC InterPro: IPR001762; Disintegrin.

CC InterPro: IPR002870; Pep_M12B_prop.

CC InterPro: IPR001590; Reprolysin.

CC InterPro: IPR000884; TSPI.

CC InterPro: IPR000130; Zn_Mtpeptdse.

CC Pfam: PF000090; tsp_1; 2.

CC Pfam: PF01421; Reprolysin; 1.

CC Pfam: PF01562; Pep_M12B_prop; 1.

CC SMART: SM00209; TSPI; 2.

CC PROSITE: PS02015; ADAM_MEPRO; 1.

CC PROSITE: PS00142; ZINC_PROTEASE; 1.

CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.

CC Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;

CC Repeat; Extracellular matrix; Heparin-binding.

CC SIGNAL 1 27

CC PROPEP 28 214

CC CHAIN 215 890

CC METAL 364 365

CC ACT_SITE 365 365

CC METAL 368 368

CC METAL 374 374

CC DOMAIN 439 526

CC DOMAIN 527 583

CC DOMAIN 584 690

CC DOMAIN 691 832

CC DOMAIN 833 890

CC DOMAIN 202 205

CC CARBOHYD 345 345

CC CARBOHYD 401 401

CC CARBOHYD 466 466

CC CARBOHYD 491 491

CC CARBOHYD 600 600

CC CONFLICT 195 195

CC CONFLICT 413 440

CC SEQUENCE 890 AA; 96671 MW; 57D70EE03D5739D3 CRC64;

Query Match 41.4%; Score 2135.5; DB 1; Length 890;

Best Local Similarity 45.3%; Pred. No. 3.9e-143;

Matches 426; Conservative 150; Mismatches 255; Indels 109; Gaps 24;

Qy 1 MLLGLTLTAF-GRTAGGEPPEVVPVIRLDPDINGRRYRWGPDSDGGLIFQ 56

Db 14 LLLLLLLLPLARGAPAPAGGAS-ELVVPTRL-----PSAGE--LALH 57

Qy 57 ITAFQEDFYLHLTDAFAPAFSTHGLVPLQGLTSSDLRRCFYSGDVNAEPDSFAA 116

Db 58 LSAFGKGVLRAPDSDFLAEPKIERLGGSGRA-TGGERGLRGCFSTGTNGEPESLAA 116

Qy 117 VSLGGLRGAGYGAEEVISPFLNAPAAQNSQAHLRLRGVGGSGDPTSCGV 176

Db 117 VSLRGLSGFLLDGEETIQP-QGAGSLAQ-----PHRLQR----- 153

Qy 177 ASGNPAILRALDPKPR---RAGFGESRSR-----RRS 207

Db 154 ---WGPAGARPL-PRGPEWEVETGEGORQGRDQHOEDSESEEEAEAGASEPPPLGAT 209

Qy 208 GRARFVSIPRYVETLVVAVDSVYKFGADLEHYLLTLATAARLYRHPISILNPINIVV 267

Db 210 SRTFRFSEARFVETLVVADASMAAFYADLQNHILTLMSVAARIYKPSIKNSINLWV 269

Qy 268 KVLRLRDSGPKVTGNAALTLRNFCAWQKLNKVSOKHPEYWDTAILFRQDLGAT-T 326

Db 270 KVLIVDEKMGPEVSDNGGLTLRNFCAWQKLNKVSOKHPEYWDTAILFRQDLGAT 329

Qy 327 CDTLGMDVGTMCOPKSCSVIEDDGLPSAFTTAHELGHVFNHMPDNVKNVCEVEFGKLA 386

Db 330 CDTLGMDVGTMCOPKSCSVIEDDGLPSAFTTAHELGHVFNHMPDNVKNVCEVEFGKLA 389

Qy 387 NHMSPTLIQIDRANPWSACSAAITDFLDSDGHDGCLLDQSPKIPISLPEDLPG--ASYTL 444

Db 390 HHVWAPLVHLNQTLPWSPCSAMVLTLLDGGHGDCLLDAPGAALPLTGLPGRMALYQL 449

Qy 445 SQCEELAFVGSKPCP---YMQYCTKLWCTCAKAGQWVCQTRH--FPWADGTSCEGKLC 499

Db 450 DQCRQIIFGPDFRHCPNTSAQDVCAQLWCHTDG-AEPLCHTKNGSLPWADGTPCGPHLC 508

Qy 500 LKGCVERHNLNHR--VDGSWAKWDVPGCSRTCGGVQVQLARRQCTNPTPANGKCYCEG 557

Db 509 SEGSCLPPEEVEVERPKPVVDGGAAPWGPWGECSTCGGVQVSHRECKDPEPQNGRYCLG 568

Qy 558 VVKYRNCNLPCPSSASGKSFREEQCEAFNGYNHSTNRLTLAVAWVPKYSGVSPDKCK 617

Db 569 RRAKYOSCHTECP--PDGKSFREQCEAFNGYNHSTNRLTLAVAWVPKYSGVSPDKCK 625

Qy 618 LICRANGTGYFVLAPKVVDTLCSPTDSVQVQCKIACGDCNGLGSKKRFDCGCGG 677

Db 626 LFCRARGSEFKVFEAKVIDGTLCGPETLAICVRGQCVKAGCDHVDSPKLDKCGCGG 685

Qy 678 DNKSKKYTGTLTKPMHGFNFVAIPAGASSIDIRQGYKGLIGDDNYLALKNSQGYLL 737

Db 686 KNSCRKYSGLSTPTNYGNDIVTIPACATNDYKQHPGVQNDGNYLALKTAGQVLL 745

Qy 738 NGHFVVSVERDLVVKSLRYSGTGTAVESLQASRPILPTVEVLSV-GKMTFPRVRY 796

Db 746 NGNLAIQAIEQDILVKYSLTILKYSIATLERLQSPRLPEPLTVQLLTVGPEVPPKVKY 805

Qy 797 SFYLPKEPREDEKSHPKDPGSPVLSLNSQVQEPDPPARWVAGSWGSPCASCG 856

Db 806 TFFVFNPD--VDFSMOSSKERATNIIQLLH-----AQWVLGDWSECSSTCG 850

Qy 857 SGLQKRAVDCRGSAGQRTVPACDAAHPRVETQAC-GEPCP 895

Db 851 ACWQRTVECDRPSGQASA-TCNKALKPEDAKPCESQLCP 889

RESULT 5

AT88_MOUSE STANDARD; PRT; 905 AA.

AC P57110;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ADAMTS-8 precursor (BC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2).

GN ADAMTS8.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_Taxid=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20079168; PubMed=10610729;

RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;

RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on

RT mouse chromosome 9 and human chromosome 11.";

Genomics 62:312-315(1999).

CC - FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES (BY SIMILARITY).

CC - COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC - SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).

CC - TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT LUNG AND HEART AND LOW EXPRESSION DURING MOUSE DEVELOPMENT.

CC - DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

CC - PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).

CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC - SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC - SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AF175282; AAF25805.1; -.

DR HSSP: P34179; 1IAG.

DR MEROPS: M12.226; -.

DR MGD: MGI:1353468; Adamts8.

DR InterPro: IPR001762; Disintegrin.

DR InterPro: IPR001590; Reprolysin.

DR InterPro: IPR000884; TSP1.

DR InterPro: IPR000130; Zn_Mtpeptdse.

DR Pfam: PF00090; tsp_1; 2.

DR Pfam: PF01421; Reprolysin; 1.

DR SMART: SM00209; TSP1; 2.

DR PROSITE: PS00142; ZINC_PROTEASE; 1.

DR PROSITE: PS02015; ADAM_MERO; 1.

DR PROSITE: PS00092; TSP1; 1.

DR PROSITE: PS00427; DISINTEGRIN_1; FALSE NEG.

DR Hydrolase: Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

KW Repeat: Extracellular matrix; Heparin-binding.

FT SIGNAL 1 28 POTENTIAL.

FT PROPEP 29 228 BY SIMILARITY.

FT CHAIN 229 905 ADAMTS-8.

FT METAL 378 378 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT_SITE 379 379 BY SIMILARITY.

FT METAL 382 382 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).

FT DOMAIN 453 541 DISINTEGRIN-LIKE.

FT DOMAIN 542 598 TSP TYPE-1 1.

FT DOMAIN 599 705 CYS-RICH.

FT DOMAIN 706 847 SPACER.

FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 905 AA; 98879 MW; 124D4132B33A0CAE CRC64;

Query Match 40.9%; Score 2113; DB 1; Length 905;

Best Local Similarity 45.5%; Pred. No. 1.5e-141;

Matches 431; Conservative 148; Mismatches 260; Indels 108; Gaps 24;

QY 2 LLLGLTL-----AFAGTAGGFEPREVVPVIRLDPDINGRRYWRGPDGQGL 53

DB 13 LLLLLLPPPLVCGAPAGPTGA--QASELVVPTL-----PGASE--L 55

QY 54 IFQITAFQEDYLLHTPDQAFLAFSTHGLGVPLQGLTGSSDLRCFTYSGDGNAPDS 113

DB 56 AFLHSAFGQGVLLAPDASFLAEFKIERLG-GSSAAGGEPGLRGCFSTGVNGERES 114

QY 114 FAASLCCGLRGAFGYRGAEEVVISPLPNASAPAAQRNSOGA-----HLLORGVPGP 166

DB 115 LAAMSCVAGSGSFLLAGEEFTIOP-----QGAGDSLDPHRLQRWG-PCQR 160

QY 167 SGDPSTSCGVASGNPAILRALDPYKPRRAGFCESRS-----RRSG----- 208

DB 161 REDP-----GLAAAEVFFLPQGLEWEVGMNGQGQERSDEEDRKQDKREGLKETSRY 216

QY 209 -----RAKRFVSIPIRYVETLVVADESVMKHGADLEHYLLTLATLARLHRSILN 260

DB 217 PPPGSTRKRFVSEARFVETLLVADASAAAFYGTDLQNHILTVMSMAARIYKHSIRN 276

QY 261 PINIVVVKVLLLRDRDSGPKVTGNAALTNRNFCAMOKKLNKVDKHPYWDTAILTRQD 320

DB 277 SVNLVVVKVILVEKERWGPESVNDGGLTLNFCNSWRRENKPSDRHPEHYDTAILTRQ 336

QY 321 LCG-ATTCDILGMADVMTCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHNVKCEE 379

DB 337 FCGEGEQCDTLGMADVGTICDPDKSCSVIKDEGLQAYTLAHELGHVLSMHPHDSKPCVR 396

QY 380 VFGKLRANHMSPTLIQIDRANPWSACSAAIITDFLDSHGDCLLDQPSKPISLPEDLP 439

DB 397 LFGPMGKYHMAFFIHNKTLNPSPCSAVYIYELDDHGDCLLDAPTSVLPPLTGLPG 456

QY 440 AS--YTLSSQCELAFGVSKPCP---YMQYCTKLWCTGKAKGMVQCQTRH--FPWADGTS 492

DB 457 HSTLYELDQCKQIFGDPDFRHCNPTSVEDICVOLCARHRDSDPEICTHKNGLWADGTP 516

QY 493 CGEGKCLKGCACVERHNLN--KRVDSGWAKWDPYPCSTCGGQVQLARRQCTNTPAN 550

DB 517 CGPHCLDGDGCVLKEDVENPKAVVDGWDGWPWPGQCSRTCGGIGQFSNRECDNMPQN 576

QY 551 GGYCEGVRYKRYSCNLEPCPSSASGSKSFRECEAFNGYNNHSTNRLTLAVAMVPKYSV 610

DB 577 GGRFCLGERVKYQSCNTECP--PNGSKSFRECEAFNGYNNH--TDLGNFLQWVPKYSV 633

QY 611 SPRDKCLICRANGTYFYVLAAPKVDGTLCSPDSTSVQVQKICIKAGCDNGLSKKRF 670

DB 634 SPRDRCKLFCRARGSEKVFYFAKVDGTLCSGPDTLCSVRCQVACGCDHVVNSPKKLD 693

QY 671 KCGVCGDGNKCKVTKGLFTKPMHGVNVAIPAGASSIDIRGKYLIGDDNYLAKN 730

DB 694 KCGVCGGKGTACRKISGTFPSFGINDIVIPAGATNIDVKRSHGPNVNDGSGYLAKT 753

QY 731 SQKYLINGHFVSAVERDLVVKGSLRYSGTGTAVESLQASRPILPTVEVLSV-GKM 789

DB 754 ANGOYLINGNLATSAEQDILVKGITLYKSGSMATLERLQSFQALPELTVQLTVSGEV 813

QY 790 TPRVRYSFLLPEPEDEKSHKDPKRGPSVLSNSVLSNQVEQDDPPPAKRWAGSWG 849

DB 814 FPPKRYTFEFPND--MDFSVQNSKERAT--NIIQSL-----PSAEWVLGDWS 858

QY 850 PCSASGSGGLQKRAVDRCGSAGQRTVPACDAHRPVEQACG-EPGP 895

DB 859 ECFSTCGSGWQRTVECRDPSGQAS-DTCDALKPEDAKPGSQPCP 904

RESULT 6

AT99 HUMAN STANDARD; PRT; 1629 AA.

ID AT99_HUMAN Q9NR29;

AC Q9P2N4; Q9NR29;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).

GN ADAMTS9 OR KIA1312.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (SHORT ISOFORM).

RC TISSUE=Fetal;

RX MEDLINE=20396138; PubMed=10936055;

RA Clark M.E., Kelner G.S., Turbeville L.A., Boyer A., Arden K.A.,

RA Maki R.A.;
 RT "ADAMTS 9, a novel member of the ADAM-TS/Metallospandin gene
 family.";
 RL Genomics 67:343-350(2000).
 RN [2]
 RP SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM).
 RC TISSUE=Brain;
 RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVI.
 The complete sequences of 150 new cDNA clones from brain which code
 for large proteins in vitro.";
 RL DNA Res. 7:65-73(2000).
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 matrix (By similarity).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES.
 CC EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG,
 COLON. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN
 PLACENTA. SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR
 THYMUS.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
 SIMILARITY).
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
 SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF261918; AAF89106.1;
 CC EMBL: AB037733; BAA92550.1;
 CC HSP: P15167; IATL.
 CC MEROPS: M12.021;
 CC Genes: HGNC:13202; ADAMTS9.
 CC MIN: 605421;
 CC InterPro: IPR001762; Disintegrin.
 CC InterPro: IPR002870; Pep_M12B_propep.
 CC InterPro: IPR001590; Reprolysin.
 CC InterPro: IPR000884; TSP1.
 CC InterPro: IPR000130; zn_M12peptdse.
 CC Pfam: PF00090; tsp_1; 11.
 CC Pfam: PF01421; Reprolysin; 1.
 CC Pfam: PF01562; Pep_M12B_propep; 1.
 CC SMART: SM00209; TSP1; 12.
 CC PROSITE: PS50215; ADAM_MEPRO; 1.
 CC PROSITE: PS00421; DISINTEGRIN_1; FALSE_NEG.
 CC PROSITE: PS50092; TSP1; 9.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Extracellular matrix; Alternative splicing.
 FT SIGNAL 1 18
 FT PROPEP 19 287
 FT CHAIN 288 1629
 FT DOMAIN 509 587
 FT DOMAIN 589 642
 FT DOMAIN 645 752
 FT DOMAIN 753 880
 FT DOMAIN 999 1053
 FT DOMAIN 1056 1108
 FT DOMAIN 1111 1156
 FT DOMAIN 1184 1239
 FT DOMAIN 1240 1295
 FT TSP TYPE-1 6.

FT	DOMAIN	1332	1383	TSP TYPE-1 7.
FT	DOMAIN	1386	1439	TSP TYPE-1 8.
FT	DOMAIN	1445	1498	TSP TYPE-1 9.
FT	DOMAIN	1501	1554	TSP TYPE-1 10.
FT	DOMAIN	1562	1612	TSP TYPE-1 11.
FT	DOMAIN	88	96	POLY-SER.
FT	SITE	223	223	CYS-TER. SWITCH (POTENTIAL).
FT	METAL	434	434	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	435	435	BY SIMILARITY.
FT	METAL	438	438	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	444	444	ZINC (CATALYTIC) (BY SIMILARITY).
FT	CARBOHYD	112	112	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	135	135	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	271	271	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	749	749	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	840	840	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1213	1213	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1267	1267	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	1064	1072	CLVTCGKGH -> VRWEGCYFP (IN SHORT ISOFORM).
FT	VARSPLIC	1073	1629	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	367	367	F -> L (IN REF. 1).
FT	CONFLICT	367	367	F -> L (IN REF. 1).
SQ	SEQUENCE	1629 AA;	182649 MW;	C1C4CEFF58B941F CRC64;
Query Match 38.1%; Score 1965; DB 1; Length 1629;				
Best Local Similarity 39.9%; Pred. No. 8.8e-131;				
Matches 411; Conservative 145; Mismatches 351; Indels 122; Gaps 21;				
QY	22	EREVYVIRLDP-----	DINGRRYYW-----	RGPEDSGQGLFIQITAFQ 61
DB	47	EYEIVSRVNALGEPFTNHFKTRRSINSATDPWPAFASSSSSTSSQAHRLSFG 106		
QY	62	EFYFLHLPDAQFLAPAPSTEHLGVP----	LOGLTGSSDLRRFCYSGDVNAEPDPAFV 117	
DB	107	QOFLNLNANAGFIAPLFTVLLGTPGVNQYKFYSEEAELKHCFKYVNTNSEHTAVI 166		
QY	118	SLCGGLRGAFYGRGAEYVVISPLPNASAPAAQRNSQAGHLLQRRGVPGSGDPTSRCOVA 177		
DB	167	SLCSGMLGTFRSHDGYFTEPLQSDMEQDEEONKPHIYRRSAPREPSTGRHACDTS 226		
QY	178	SGWNPAILRALDPYKPRAGFGE-----	SRSSRRSG 208	
DB	227	EHKN---RHSKDKKTRARKWGERINLAGDVAAALNSGLATEAFAYGNKNTDREKTRHR 283		
QY	209	RAKRFVSIPIRYVETLVVADESMVKFEGADLEHYLLTLATARLYRHPHSILNPIVVKV 268		
DB	284	RTKRFLSYPRFVEVLVADNRVSYHGENLQHYLTLSIVASIVKDFPSIGNLINIVN 343		
QY	269	VLLLRDRDSGPKVTGNAALTLRNFCWQKLNKVSCKHPEYWDTAILTRQDLCA-TTC 327		
DB	344	LIVIHNEQDQPSISFNAQTTLKNFCQWQHSKNSPGGIH---HDTAVLLTRQDICRAHDKC 400		
QY	328	DTLGMDVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRAN 387		
DB	401	DTLGLAELGTICDPYRSCSISDSGLSTAFITAHELGHVFNPHDNNDNKKCEE-GVKSPQ 459		
QY	388	HMSPTLIQIDRANPWSACSAIITDFLDHSGDCLLDQP-SKPISLPEDLPASVYLSQ 446		
DB	460	HVNAPTLNFTNPMWMSKSRKYITEFLDTGYCECLLNPEPSRPYLPVQLPGILYNNVK 519		
QY	447	QCELAFGVSGKPCPYMOYCTKLWC---TGKAGQWVCQTHPEHPWADGTSCGSGKLCCKGA 503		
DB	520	QCELIFGSGSQVCPYMMQCRRLMNNVNGVHKG---CRTQHTFWADGTCECEPKCKYGF 576		
QY	504	CVERHNLNKHVRDGSNAKWDPYGPCSRTCGGGVQLARRQCTNPTTPANGKYCEGVVRKYR 563		
DB	577	CVPEK-EMDVPVTDGSGWSGSPFTCSRTCGGKTKTARECNRPENKNGKCYCVGRMKFK 635		
QY	564	SCNLEPCPSSASGSKSPREEQCEAFNGYHNSTNRLTLAVAWVPKYSGVSPDRCKKLCIRAN 623		
DB	636	SCNTEPCLKQK--RDFRDEQCAHFDGKHFNINGLLPNRWVPKYSGILMKDRCKLCFCRYA 693		
QY	624	GTGYFYVLAIPKVVVDGTLCSPDSTSVCVQKCIKAGCDGNLGSKKRFDKCGVCGDGNKSKC 683		

```

Db 694 GNTAYQLRDRVIDGTPCGQDNDICVQLCRQAGCDHVLNSKARRDKCGVCGDSSCK 753
QY 684 KYTGFTKPMHGYNFWAIPAGASIDTRQRYGKGLIGDDNYLALKNSQGYLNGHFVV 743
Db 754 TVAGFTNVHGYNTVVRIPAGANIDVRQHSFGSETDDDDNYLALSSSGEFLNGFVV 813
QY 744 SAVERDLVVKGLSLRYSTGTAVESLQASRPILPTVLEVLVSGKMTPPRVRYSYFLPK 803
Db 814 TMAKREIRIGNAVVEYSSETAVERINSTRIEQLLOLVLSVGLNPDVRYSFNPIE 873
QY 804 PREDK---SSHPKDRGP-----SVLHNSVLSSNQVEQDPPDRPP--- 840
Db 874 DKPOQFYWNH-----GPWQACSPQCGERRKRLVCTRESQDQVSV---DQRCRLPQ 925
QY 841 -----ARVAGSWGSPCSACSGSLQKRAVDCR-----GSAGQRTVPACDAHR 883
Db 926 HITEPCGTDCDLRHVAVSRSCSQCGGLGYRTLDIYCAKYSRLDGKTEKVDGFCSSHPK 985
QY 884 PVETQACGECPT--WELSAWSPCKSCGGRGFRRLKCVGHGGRLLARDQCNLHRKPQE 941
Db 986 PSNREKCSGECNTGWRYSAWTECSKSCDGTGTRRAICVNTNRDVLDDSKCT-HQEKVT 1044
QY 942 LDFCVLRPC 950
Db 1045 IORCSEFFC 1053

RESULT 7
ATSA_HUMAN
AC 075173; Q9UN83; STANDARD; PRT; 837 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
DE (ADMP-1).
GN ADAMTS4 OR KIAA0688.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99286303; PubMed=10356395;
RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
RA Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
RA Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,
RA Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
RA Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
RA Trzaskos J.M., Arner E.C.;
RT "Purification and cloning of aggrecanase-1: a member of the ADAMTS
RT family of proteins.";
RL Science 284:1664-1666(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX Sawai I., Nagase H., Saklatvala J., Clark A.R.;
RT "ADAMTS-4 genomic locus.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=20400518; PubMed=10827174;

```

```

RA Tortorella M.D., Pratta M.A., Liu R.-O., Abbaszade I., Ross H.,
RA Burn T.C., Arner E.C.;
RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
RT aggrecan substrate recognition and cleavage.";
RL J. Biol. Chem. 275:25791-25797(2000).
CC -!- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE
CC CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN
CC ALZHEIMER'S DISEASE.
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu|-Ala-393
CC site.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED
CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
CC -!- INDUCTION: BY INTERLEUKIN-1.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO AS ADAMTS2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB014588; BAA31663.1; -
DR EMBL; AF148213; AAD41494.1; -
DR EMBL; AY044847; AAL02262.1; -
DR MEROPS; M12.221; -
DR Genew; HGNC:220; ADAMTS4.
DR MIM; 603876; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_Mtpeptdse.
DR Pfam; PF00090; tsp_1; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSP1; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Extracellular matrix.
FT SIGNAL 1 51
FT PROPEP 52 212
FT CHAIN 213 837
FT SITE 194 194
FT METAL 361 361
FT ACT_SITE 362 362
FT METAL 365 365
FT METAL 371 371
FT DOMAIN 437 519
FT DOMAIN 520 576
FT DOMAIN 577 685
FT DOMAIN 686 837
FT DOMAIN 247 252
FT DOMAIN 68 68
FT CARBOHYD 77 77
FT CONFLICT 626 626
FT CONFLICT 682 682
FT SEQUENCE 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;
SQ

```

Query Match

37.1%; Score 1917; DB 1; Length 837;

Best Local Similarity 46.3%; Pred. No. 9, 7e-128;

Matches 381; Conservative 123; Mismatches 267; Indels 52; Gaps 16;

```
QY 1 MLLGLTILTAFAAGTGGPEPEVVEVPIRLDPDINGRYVWRGPDSDQGLIFQITAF 60
Db 37 LLLLLLASLPASRLAPRIPEEIVFPKLGNSVL-----PQSGAPARLLCRLOAF 88
QY 61 QEDFYHLTPDAFLAPAFSTEHLGVLPLQGLTGGSDLRRCFYSGDVNAEPDSFAAVSVC 120
Db 89 GEYLLLEODSGVQVEGLTVQVIGQAPE-LLGAEPP--GTVLTGTINGDPESVASLHWD 145
QY 121 GG-LRGAFGRGAYEIVISPLPNASAPAAQNSOGAHLORRGVPGPSDPTSRGCVASG 179
Db 146 GGALLGVQYRGAEHLQPLEGGTTPNSA--GGPGAHILRRK-----SPASGQGPNCNV--- 196
QY 180 WNPAILRALDPYKPRAGFCESRRSRGRKRFVSIPRYVETLVVADSMVKFHHGADLE 239
Db 197 -----KAPLGSPSPRR--RAKRFASLSRFVETLVVADDDKMAAFHGAGLK 239
QY 240 HYLTLTATARYRHPSILNPINIVVVKVLLLRDSGSKVVTGNALTLRNFCAWKKL 299
Db 240 RYLLTVMAAAKAFKHPISIRNPVSLVTRVLVILGSGEGPQVGPSSAAQTLRSECAWQRL 299
QY 300 NKYSDKHPEYWDTAIFTRDLCGATTCDFLGMADVTGMDPKRSCSVTDDGLPSAFTT 359
Db 300 NTPEDSDPHDFTAIFTRDLCGVSTCDFLGMADVTGCDPARSCAIVEDDGLQSAFTA 359
QY 360 AHELGHVFNPHNPVKNVCEEVFGKL-RANHMSPTLIQIDRANPWSACSAIITDLDG 418
Db 360 AHELGHVFNHMLHNSKPCISLNGPLTSRHMVAPVMAHVDPEPWPSCSAREITDLDNG 419
QY 419 HGDCLLDQPSKPLSPEDLPASVYLSQCELAFGVGSKPCPYM-QYCTKLVCTGKAKGQ 477
Db 420 YGHCLDKPEAPLHLPVTFPKDYADROQQLTFGPDSTRHCPQLPPPCAALWCSHNLGH 479
QY 478 MVOCTREPPNADGSCGEGKCLKGACVERHNLKHKVD--GSWAKWDYPGCSRTCGG 535
Db 480 AMCQTKHSPWADGTPCPAOCACGRCRLHMDQLQDFNIPOAGGNGPWPWGDCSRCTCGG 539
QY 536 VQLARRCQNTPTPANGKYGCEGVKRVYRSCNLEPCPSASGKSFREPOCEAFNGYNHSTN 595
Db 540 VQFSRDCRTPVPRNGKYGCEGRTRFRSCNTEPCP-TGSALTFRBQCAA---YHNRTD 595
QY 596 RLTL---AVAWPKYSGVSPDKCKLCIRANGTCYFVVLAPKVVDGTLCSPDSTSVCVQG 652
Db 596 LFKSPGPMWVRYTGVAPQDQCKLQARALGYVYVLEPRVVDGTPCSPDSSSVCVQG 655
QY 653 KCIRAGCDNGLSKRRDKCVCQGDKNKCKKVTGLFTKPMHGNFYVVAIPAGASSIDIR 712
Db 656 RCIHAGCDRIIGSKKFKDKCMVCGDGGSGSKQSGSEFKFRYGYNNVVTIPAGATHLVR 715
QY 713 QRGYKGLIGDDNYLALKNSQKYLNLNGHFVVSAYERDLWVKGSL-LRYSGTGTAVESLQA 771
Db 716 QQGNPG--HRSIYLALKLPDGSYALNGEYTLMPSTVDVLPVGAVSLRYSGATASETLSG 773
QY 772 SRPILPPLTVLSVGKMTTPRVRYSYFLPKPEPREDSSHPKD 814
Db 774 HGPLAQPLTLQVLVAGNQDTRLRYSFVFPRTPTPTPDQ 816
```

RESULT 8

```
ATSS_MOUSE STANDARD; PRT; 930 AA.
ID ATSS_MOUSE
AC Q9R001;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
DE (ADMP-2) (Implantin).
GN ADAMTS5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX SEQUENCE FROM N.A.
RA MEDLINE=9395124; PubMed=10464288;
RA Hurskainen T.L.; Hirohata S.; Seldin M.F.; Apte S.S.;
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RT zinc metalloproteases."
RL J. Biol. Chem. 274:25555-25563(1999).
CC -|- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
CC PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC -|- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393
CC site.
CC -|- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -|- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PERI-
CC IMPLANTATION PERIOD IN EMBRYO AND TROPHOBLAST AND AT LOW OR
CC UNDETECTABLE LEVEL THEREAFTER.
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -|- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF140673; AAD56356.1; -.
DR MEROPS: M12.225; -.
DR MGD; MGI:1346321; Adamts5.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Repolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF00090; tsp_1; 2.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPI; 2.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 261 POTENTIAL.
FT CHAIN 262 930 ADAMTS-5.
FT SITE 209 209 CYSTEINE SWITCH (POTENTIAL).
FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 411 411 BY SIMILARITY.
FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 485 566 DISINTEGRIN-LIKE.
FT DOMAIN 567 623 TSP TYPE-1 1.
FT DOMAIN 624 731 SPACER.
FT DOMAIN 732 874 CYS-RICH.
FT DOMAIN 875 930 TSP TYPE-1 2.
FT DOMAIN 41 46 POLY-ALA.
FT DOMAIN 257 261 POLY-ARG.
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
```


RESULT 9
ATS5_HUM
ID ATS
AC Q9U
DT 16-

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF142099; RAD49577.1; -
 CC EMBL; AF001698; BAA95504.1; -
 CC EMBL; AF001697; BAA95503.1; -
 CC EMBL; AF141293; AAF02493.1; -
 CC HSP; Q9FW35; 1BUD.

DR MEROPS; M12.225; -
 DR Genew; HGNC:221; ADAMTS5.

DR InterPro; IPR001762; Disintegrin.
 DR InterPro; IPR002870; Pep_M12B_propep.

DR InterPro; IPR001590; Reprolysin.
 DR InterPro; IPR000884; TSPL.

DR InterPro; IPR000130; Zn_MTPeptide.
 DR Pfam; PF00090; tsp_1; 2.
 DR Pfam; PF01421; Reprolysin; 1.

DR SMART; SM00209; TSPL; 2.
 DR PROSITE; PS00215; ADAM_MEPRO; 1.
 DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.

DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

KW Repeat; Extracellular matrix.
 FT SIGNAL 1 16 POTENTIAL.

FT PROPEP 17 261 POTENTIAL.
 FT CHAIN 262 930 ADAMTS-5.

FT SITE 209 209 Cysteine SWITCH (POTENTIAL).
 FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT_SITE 411 411 BY SIMILARITY.
 FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 485 566 DISINTEGRIN-LIKE.

FT DOMAIN 567 623 TSP TYPE-1 1.
 FT DOMAIN 624 731 SPACER.

FT DOMAIN 732 874 TSP TYPE-1 2.
 FT DOMAIN 875 930 TSP TYPE-1 1.

FT DOMAIN 37 41 POLY-ALA.
 FT DOMAIN 257 261 POLY-ARG.

FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 138 138 A -> G (IN REF. 2).
 FT CONFLICT 614 614 R -> H (IN REF. 3).

FT CONFLICT 692 692 P -> L (IN REF. 2).
 SQ SEQUENCE 930 AA; B64281502F28193B CRO64;

Query Match 35.8%; Score 1847.5; DB 1; Length 930;
 Best Local Similarity 44.0%; Pred. No. 8.9e-123;

Matches 392; Conservative 113; Mismatches 291; Indels 95; Gaps 22;
 Qy 48 SGDQGLIFQITAFQEDFYHLTPDAQFAPAFSTHGVLPQLGTLGSSDLR---RCFYS 104

Db 82 SGGKGVLYVAGGRRELLDERDGSVGIAGF-----VPAGG--GTSAPWRHSHCFYR 133

Qy 105 GDVNAEPDSFPAVSLCGLGCAFYCAEYVISP-----PNASAPAAQNSQGA----HLL 157

Db 134 GTVDASPRSLAVFDLCGLDGLGFFAVKHARYTLKPLRGPAWEKEKGVYGDGSRILHVV 193

Qy 158 QRRG-----VPGGPS-----GDPTSRGCVASGWNPAIIL--RALDPYKP 193

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Db 300 RYSHASINIRLAVKVVVGLGDKSLEKSKNAATLKNFCKWQHNLGDDHEHY 359

Qy 311 DTALFTRDQLCGATTCDTLGMADVGMCDPKRSCSVIEDGLPSAFTTAHELGHVFNMP 370

Db 360 DAAILFTREDLGGHSCDTLGMADVGTICSPERSCAVIEDDGLHAFTVAHEHLLGLS 419

Qy 371 HDNVKVEEVFGKLRANHMSPFTLIQIDRANPWSACSNAITDLDSCGHGCLLDQSKP 430

Db 420 HDDSKFEETFGSTEDKRLMSILTSIDASKPWSKTSATITEFLDDGGHCLLDPRKK 479

Qy 431 ISLPEDLPASVTLSSQOCELAFGVGSKPCPYMQVCTKLMCTGKAKGQMVCCOTRHFPPWADG 490

Db 480 ILGPEELFGQYDATQCQNLFTGPEYSVCPGMDVCAWLCAVVRGQGMVCLTKLPAVEG 539

Qy 491 TSCGEGKILKGACVER-----HNLNKHVRVDGSMKWDPPYGPCSRTCGGGVQLARRCTN 545

Db 540 TPCGKGRICLOGKVDKTKKKVYSTSSH---GNWGSWGSWGQCSRSCGGGVQFAYRHNN 596

Qy 546 PTPANGKYGCEGVVKYRSNLEPCPSSASGSKSPREEOCEAFNGYNHSTNRLTLAVAWVP 605

Db 597 PPRNNGRYCTGKRAIYRSCSLMPCP--PNGKSPRHEQCEAKNGYQSDAKGVKTFVEWVP 654

Qy 606 KYSGVSPRDKCKLICRANGTGVYVLAAPKVVVDGTLCSDDSTSVCGQKICAGCDGNLGS 665

Db 655 KYAGVLPADVCKLTCAKGTGYVVFSPKVDGTGTECRYSNSVCRGKCVGTGCGGIIGS 714

Qy 666 KKRFDKCGVCGGDNKSCCKVTGLFTKPMHGNFVVAIPAGASSIDIRQRYKGLIGDDNY 725

Db 715 KLIQDKCGVCGGDNSSCTKIIVGTFNKKSKGYTDVVRIPEGATHIKVRQKAKDQTRFTAY 774

Qy 726 LALKNSQKYLNGHFVVSVERDLVVGSLRLRYSGTGTAVESL-----QASRPILPLELT 780

Db 775 LALKKNGEYLNGYKIMISTETIIDINGTVNYSWGSVSHRDDFLHMGYSATKEI---LI 831

Qy 781 VEVLSVGMKTPRVYRYSFVLPKPEPKEDKSSHPKDPGRPSVLHNSVLSL-SNQVEQPDPR 839

Db 832 VOILATDPTKPLDVYRYSFVFPK-----STPK-----VNSVTHSGSNKVSHTSQP 877

Qy 840 PARWAGSWGPCASGSLQKRAVDRGSAQORTVPACDAHRAHREVTQAC 890

Db 878 --QWVTGPTMACSRCTDGTGHTRTVQCQ--DGNRKLAKGCLPSQRPSPAFKQC 925

RESULT 10

ATSA4_RAT

ID ATSA4_RAT STANDARD; PRT; 630 AA.

AC Q9ESP7; Q9ESP8; Q9ESP6;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ADAMTS-4 precursor (BC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)

DE (Fragment).

GN ADAMTS4.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_Taxid=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-Wistar; TISSUE=Brain;

RX MEDLINE=20415831; PubMed=10961658;

RA Sato K., Suzuki N., Yokota H.;

RT "ADAMTS-4 (a disintegrin and metalloproteinase with thrombospondin motifs) is transcriptionally induced in beta-amyloid treated rat astrocytes.";

RL Neurosci. Lett. 289:177-180(2000).

CC -!- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES.

CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393 site.

```

CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -!- TISSUE SPECIFICITY: ONLY EXPRESSED IN BRAIN.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB042272; BAB16474.1; -
DR EMBL; AB042271; BAB16473.1; -
DR EMBL; AB042273; BAB16475.1; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_Mtpeptidse.
DR Pfam; PF00090; tsp_1; 2.
DR Pfam; PF01421; Reprolysin; 1.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSP1; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00427; DISINTEGRIN.1; FALSE_NEG.
DR Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
KW Extracellular matrix.
KW NON_PEP 1
KW PROPEP 1
KW CHAIN 6
KW METAL 154
KW ACT_SITE 155
KW METAL 158
KW METAL 164
KW DOMAIN 233
KW DOMAIN 316
KW DOMAIN 367
KW DOMAIN 368
KW DOMAIN 479
KW DOMAIN 490
KW CARBOHYD 96
KW CARBOHYD 474
KW SEQUENCE 630 AA; 68384 MW; 63A428753167C7EF CRC64;
KW -----
KW BY SIMILARITY.
KW ADAMTS-4.
KW ZINC (CATALYTIC) (BY SIMILARITY).
KW BY SIMILARITY.
KW ZINC (CATALYTIC) (BY SIMILARITY).
KW ZINC (CATALYTIC) (BY SIMILARITY).
KW DISINTEGRIN-LIKE.
KW TSP TYPE-1.
KW CIS-RICH.
KW SPACER.
KW POLY-ALA.
KW N-LINKED (GLCNAC. .) (POTENTIAL).
KW N-LINKED (GLCNAC. .) (POTENTIAL).
KW MW: 63A428753167C7EF CRC64;
KW SEQUENCE 630 AA; 68384 MW; 63A428753167C7EF CRC64;
KW -----
Query Match 33.8%; Score 1745.5; DB 1; Length 630;
Best Local Similarity 52.6%; Pred. No. 8.9e-116;
Matches 325; Conservative 88; Mismatches 184; Indels 21; Gaps 9;
QY 209 RAKRFYSIPRYVETLVVADESVMKFGHADLEHYLLTLATAARLYRHPISILNPINIVVYK 268
DB 1 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2 RTRKFASLSRFVETLVVADDKMAAFHGAGLKHYYLLTVMAAAAKAFKHPISRNPNVLVWTR 61
QY 269 VLLLRDRDSGPKVTGNAALTNRNFCWAKKLNKYSDKHPEYDWTAILFTRODLGCAATCD 328
DB 1 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 62 LVLLGSGEQVPGVGSAAQLRSTFWKGLNPNPNDSDPDHFDLFTRODLGCGVSTCD 121
QY 329 TLGMADYGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCSEVEFGK-LRAN 387
DB 1 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 122 ALGMAGVGTVCDDPARSCAIVEDGLQSAFTAAHELGHVFNMLHDNSKPCANLNGQSSSR 181
QY 388 HMQSPPTLIQIDRANPHSACSAITLDFLDGSHGCDLDDPSKPLSPEDLPASYTLSSQ 447
DB 1 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 182 HVAPVMAHVDDPEEPWSPCSARFITDFLDNGYGHCLLDKPEAPLHLPEVTFPGKDYDADQ 241
QY 448 CELAFYGSKSPCYM-QYCTKLWCTGKARKGMVQCOTRHFPMWADGTCSCGKGLKGACVE 506
DB 1 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
DR Pfam: PF00090; tsp_1; 5.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR SMART: SM00209; TSP1; 5.
DR PROSITE: PS00215; ADAM_MEPRO; 1.
DR PROSITE: PS00092; TSP1; 2.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
DR Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT FT. NON_TER 1
FT PROPEP <1 207 BY SIMILARITY.
FT CHAIN 208 1077 ADAMTS-10.
FT METAL 366 366 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 367 367 BY SIMILARITY.
FT METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 376 376 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 434 520 DISINTEGRIN-LIKE.
FT DOMAIN 578 679 CYS-RICH.
FT DOMAIN 580 802 SPACER.
FT DOMAIN 521 577 TSP TYPE-1 1.
FT DOMAIN 799 860 TSP TYPE-1 2.
FT DOMAIN 862 918 TSP TYPE-1 3.
FT DOMAIN 922 976 TSP TYPE-1 4.
FT DOMAIN 981 1031 TSP TYPE-1 5.
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 714 714 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 769 769 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 866 866 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1077 AA; 118072 MW; 3914DE18DCB5F587 CRC64;

Query Match 25.1%; Score 1296; DB 1; Length 1077;
Best Local Similarity 35.4%; Pred. No. 9.4e-94;
Matches 335; Conservative 112; Mismatches 376; Indels 123; Gaps 32;

QY 24 EVVVPRIIDPPDINGR-----RYTWGPEDSGDGLIFQITAFQEDFYHLTPDAQF 74
DB 14 EIAFPTRVDH--NGALLAFSPPPRRQRGTGATAESRLFYKVASPTHELLNLTSSRL 71
QY 75 LAPAFSTHGLVPLQGLTGGSSDLRCFYSDVNAE--PDFAAVSLCGGLRGAFYRGAE 133
DB 72 LAGHVSVEYW--TREGLAWQRAARPHCLYAGHLQQAQSSSHVAISTCGGLHGLIVADEE 129
QY 134 YVISPLPNASAPAAQRNSOGAHLQRGVPGCGSDPTSRGV--ASGW--NPAILRALD 189
DB 130 YLIEPLHCGPGKSRSPESGPHVYKRSLSRPHLD--TAGGVDEKPKRPMWLTXL 187
QY 190 PYKPRRAGGESSRRSRRGRKFRVPIRYVETLIVADESMVKPHG--ADLEHYLLTLAT 248
DB 188 P--PPARPLGNETERGQPG--LKRSVSRYRYVETLIVADKMMVAYHGRDRVQYVLAIMI 244
QY 249 EARLYRHSILNPINIVVVKLLLRDSDGPKVTGNAALTLRNFCAQKKL-----NK 301
DB 245 VAKLFQDSSLGSTNVLITRILITLIDOPTLEITHHAGKSLDSFCWKOKSVNHSHGNA 304
QY 302 VSDKHPEYWDYALITFRDLG--GATTCDTLGMADVGTMTCDPKRSCSVIEDGLPSAFT 359
DB 305 IPENVANHDYAVLTIRYDICIYKNKPGTGLAPVPGMGERSCSVNEDIGLPQAFIT 364
QY 360 AHELGHVFNMPHDNVKVEEVEFG----KLRAHMMSPFLIIDLANP--WSACSAALITD 413
DB 365 AHEIGHTFGMNHGVDGNSCGARGQDPAPKLAHAHITMTK-----NPFVWSSCNRDITS 417
QY 414 FLDSGHGCCLLDQPSK--PISLPEDLPASYYTLISOCELAFGVGSKPQPMYCYCKLWCTG 472
DB 418 FLDLSGLGLCLNRRPQDFVPTVAPGQAYDADEQCRHQHGVKRSQCKYGEVCSLWCLS 477
QY 473 KAKGMVQCTRHFPWADGTSCEGKL-----CLKGACVERHNLNKHVDGSKWAKWDYGPC 528
DB 478 KSNR---CITNSIPAAEGTLQQTHTIDKGWCYKRCVDP--FGSRPEGVGDGAWGPWTGDC 533

529 SRTCGGGVQLARRQCTNPTPANGKYGCVRVKRSCHNLEPCPSASGKSFPREOCEAFN 588
534 SRTCGGGVSSSRHCDSPRTIGGKYCLGERRRHRSCTNTDCCPGS--QDFREVQCSEFD 591
589 G-----YNHSTNRLTLAVAVPKYSGVSPDKCKLLCRANGTCGYFVLAPKVVDDGTLG 641
592 SIPFRGKFKYKWKTYR-----GGGVK---ACSLTSLAEGFNFYTERAAVVDGTPC 638
642 SPDSTSVCVGKCIKAGDGNLGSKKRDKCGVCGGCKNKKCKVTGLFT--KPMHGYNFV 699
639 RPDVTDICVSGECKHVGCDRLGDLREDKRCVCGDGSACETIEGVSPASPGAGYEDV 698
700 VAIPAGASSIDIRQRYGKGLIGDDNYLALKNSQGYLLNGHFVVSADVERDLVWKSLLRY 759
699 VWPKGSVHIFIQD-----LNLSSLHALKGDOESLLEGLPGTQPQHR-LPLAGTTTQL 752
760 STGTAVESLQASRILEPLTVEVLSVGKMTPPRYYSFYLPKEPREDKSSHPRGPGS 819
753 RQGPDOVQSLEALGPINASLIYVNLARTEL--PALRYRFNAP----- 792
820 VLHNSVLSLSNOVEQPDPRPPARVWAGWPCSGSGGLQKRAVDCRGSAGQRTV--PAC 878
793 -----IARDSLPYSHYAPWKCSAQACAGSQVQAVECRNQLDSSAVAPHY 839
879 DAAHR--PVETOACG-EPC-PTWELSAWSPCKSCGRGFRRLSKC 920
840 CSAHSLPKRQRACNTEPCPPDWVVGNNSLCSRSCDAGVRSRVVC 885

RESULT 12
AT12_HUMAN STANDARD; PRT; 1593 AA.
ID AT12_HUMAN
AC P58397;
DC 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-
DE TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Petal lung;
RX MEDLINE=21264577; PubMed=11279086;
RA Cal S., Argueelles J.M., Fernandez P.L., Lopez-Otin C.;
RT "Identification, characterization, and intracellular processing of
RT ADAM-TS12, a novel human disintegrin with a complex structural
RT organization involving multiple thrombospondin-1 repeats.;"
RL J. Biol. Chem. 276:17932-17940(2001).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
CC -1- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
CC DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AJ250725; CAC20419.1; --
Genew; HGNC:14605; ADAMTS12.

MIM: 606184; --

InterPro; IPR001762; Disintegrin.

InterPro; IPR002870; Pep_M12B_propep.

InterPro; IPR001590; Reprolysin.

InterPro; IPR000884; TSPI.

InterPro; IPR000130; Zn_MTPeptidase.

Pfam; PF00090; tsp.1; 6_MTPeptidase.

Pfam; PF01421; Reprolysin; 1.

Pfam; PF01562; Pep_M12B_propep; 1.

SMART; SM00209; TSPI; 8.

PROSITE; PS0215; ADAM_MEPPO; 1.

PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.

PROSITE; PS0092; TSPI; 2.

PROSITE; PS00142; ZINC_PROTEASE; 1.

KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

KW Repeat; Extracellular matrix.

FT SIGNAL 1 25 POTENTIAL.

FT PROPEP 26 240 BY SIMILARITY.

FT CHAIN 241 1593 ADAMTS-12.

FT DOMAIN 465 544 DISINTEGRIN-LIKE.

FT DOMAIN 545 596 TSP TYPE-1 1.

FT DOMAIN 597 700 CYS-RICH.

FT DOMAIN 701 826 SPACER 1.

FT DOMAIN 827 881 TSP TYPE-1 2.

FT DOMAIN 886 943 TSP TYPE-1 3.

FT DOMAIN 947 995 TSP TYPE-1 4.

FT DOMAIN 996 1315 SPACER 2.

FT DOMAIN 1316 1364 TSP TYPE-1 5.

FT DOMAIN 1367 1423 TSP TYPE-1 6.

FT DOMAIN 1426 1471 TSP TYPE-1 7.

FT DOMAIN 1471 1471 TSP TYPE-1 8.

FT DOMAIN 302 305 POLY-GLU.

FT SITE 208 208 CYSTEINE SWITCH (POTENTIAL).

FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT_SITE 393 393 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).

FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 685 685 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 951 951 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1275 1275 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1300 1300 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1320 1320 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1503 1503 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT SEQUENCE 1593 AA; 177545 MW; 07F9F48B63BD8A3 CRC64;

Query Match

Best Local Similarity 24.7%; Score 1275; DB 1; Length 1593;

Matches 330; Conservative 155; Mismatches 374; Indels 158; Gaps 37;

QY 21 PEREVVPIRLD-----PDINGRRYWRGPESGGDGLIFQITAFQEDPYLHL 68

Db 50 PEYHVGVPRVDASGHFLSYGLHYPTSSRR---KRDLDGSEDWVYRISHEEKLFNNL 106

QY 69 TPDQAQFLAPAFSTHGLVPLQGLTGSSDLRRCFYSDYNAEPD--SFAVSLCGGLRGA 126

Db 107 TVNQGLSLNSYIMEKRYGNLSHVYKMASSAPLCHLSLGTVLQOQTRVGTAAALSACHGLTGF 166

QY 127 FGYRGAEYVISPLPNASAPAAQRNSQGAHLQRQVPGPGSDPTSRGCVASGWPAILR 186
Db 167 FQLPHGDFTIEPV--KKHPLVEGGYHPIVYRRQKPV--ETREPT--CGLKDSVNSIQKQ 220
QY 187 ALDPYKPRAGFCESRRSRSGRAKRFVSIPIRYVETLVVADSMVKFHGAD--LEHYLLTL 245
Db 221 ELWREKWERHNL--PSNLSR-----RSISKERHWETLVVADTKMIEYHGSSEVSIYITI 274
QY 246 LATAARLYRHPSTILNP INIVVVKVLLLRDRSGPKVTGNAALTLLRNFCAMQKKNKVDK 305
Db 275 MNMVTGLFHNPSIGNAIHIVVRLILLEEEOGLAIVHAEKTLSSFCWKQKWSINPKSDL 334
QY 306 HPEYWDTAILFTRODLGA--TTCDTLGMADVGTWCDPKRSCSVIEDDGLPFAFTAHHEL 363
Db 335 NPVHHDVAVLLTRKIDICAGFNPCETGLSHLSGMQCPHRSNCINIEDSLPLAFTIAHEL 394
QY 364 GHVFNMPHDNVK--VCEEVFGKLRANHMSPPTLIQIDRANPWSACSAIIITDLDGSHGDC 422
Db 395 GHSFGIHDGKENDCEPVG--RHPYIMSRQLOYDPTPLTWKSCSEYITRELDKRGWFC 451
QY 423 LLDQPSKP--ISLPEDLPGASYTLSSQCELAFGVSGKPCPYMO--YCTKLWCTGKAKGQWVC 480
Db 452 LDDIPKKGLKSKVIAPGVYIDVHHQCOLOQYGNATFCQEVENVQCOTLWCSVKG---FC 507
QY 481 QTRHFPWADGTSQGEKGLCKACVVERHNLK--HRVDGSAKWADWPGPCSRTCCGGVOL 538
Db 508 RSKLDAADGTCQGEKKKCMAGKCI---TVGKKPESIPGGWGRWSPWSHCSRTCCGQVS 564
QY 539 ARROCTNPTPANGKYCEGVRYKRSCLNEPCPSSASGKSPREEOCEAFNGYNHSTNRLT 598
Db 565 AERLCNNPEPKFGKYCTGERKRYRLCNVHPCSEA--PTFRQMCSEFDVTPYK--NEL- 620
QY 599 LAVAVPKYSGVSPRDKKLCICRANGTYFYVLABVKVVDGLC---SPDSTSVCVQKCIK 656
Db 621 --YHWFPIF--NPAHPCELYCRPIDGQFSEKMLDAVIDGTFCFEGGNSRNVCIINGICKM 675
QY 657 AGCDGNLGSKKRDKCGVCGGDKSKCKVTGLF--TKPMHGYNFVVAIPAGASSIDIRQSG 715
Db 676 VCGDEYDINATEDKCGVLCGDSGSCQTVRRMKFQESGVVDIGLIPKGARDIKVME-- 733
QY 716 YKLGIDDNLYLAKNSQ--GKYLINGHVVYSAVERDLVYKGLSLRYSGTGTAVESLQASRP 774
Db 734 ---IEGAGNFLAIRSEDEPKYVINGGFIQ--WNGNYKLAVTFQYDRKGD--LEKLMATGP 788
QY 775 ILEPLTVEVLSVGKMTPPRVRSFYLPKEPREDKSHKDPGRPSVLHNSVLSNQVEQ 834
Db 789 TWNESVMIQLLF--QVTNFGIRYEYTIQKD-----GLDNDVEQ 823
QY 835 -----PDDRP-----PDRP-----PA 841
Db 824 MYFWQYGHWTESVTCGTGIRQTAHCIKKGRGWKATFCDPQPNRGKQKCKHEKACPP 883
QY 842 RRVAGSWGPCSASCG--SGLQKRAVDCRGS--AGQRTVPACDAAH--RPVETQACGEP--C 894
Db 884 RWWAGEWEACSATCPGHEKKRTVLCIQTWVSDEQALPPTDQHLKPKTLLSCNRDILC 943
QY 895 PT--WELSAWSPCSKSCGGRGFRSRLKCVGHGRLRLARDQCNLHRKPQELDFCVLRPC 950
Db 944 PSDVTWVGNWSECSVSCGGGVRSIRSVYCAKNHD-----EPCDVTRKPNRSRALCGLQOC 995

RESULT 13

AT52_HUMAN

ID AT52_HUMAN STANDARD; PRT; 1211 AA.

AC O95450;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and

DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)

DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I

DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNP1)

(procollagen I/II amino-propeptide processing enzyme).

ADAMTS2 OR PCINP OR PCPNT.

Homo sapiens (human).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;

[1]

SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI), AND DISEASE.

TISSUE=Skin;

MEDLINE=99347935; PubMed=10417273;

Collige A., Sieron A.L., Li S.-W., Schwarze U., Petty E., Wartecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W., Byers P.H., Lapiere C.M., Prockop D.J., Nussgens B.V.; "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis are caused by mutations in the procollagen I N-proteinase gene."; Am. J. Hum. Genet. 65:308-317(1999).

-I- FUNCTION: Cleaves the propeptides of type I and II collagen prior to fibril assembly, does not act on type III collagen. May also play a role in development that is independent of its role in collagen biosynthesis.

-I- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain alpha-1(I) at Pro-|-Gln and of alpha-1(II) and alpha-2(I) chains at Ala-|-Gln.

-I- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

-I- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO COLLAGEN TYPE XIV (BY SIMILARITY).

-I- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (by similarity).

-I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI; ARE PRODUCED BY ALTERNATIVE SPLICING.

-I- PROCOLLAGEN PEPTIDASE ACTIVITY.

-I- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.

-I- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

-I- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).

-I- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos syndrome type VIIC (EDS-VIIC), a recessively inherited connective-tissue disorder characterized clinically by severe skin fragility and joint hypermobility and biochemically by the presence in skin of procollagen incompletely processed at the N-terminus.

-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

-I- SIMILARITY: CONTAINS 4 DISINTEGRIN-LIKE DOMAIN.

-I- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAINS.

-I- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

ENBL: AJ003125; CAA05880.1; --

DR MEROPS; M12.301; --

DR Genew; HGNC:218; ADAMTS2.

DR MIM; 604539; --

DR MIM; 225410; --

DR InterPro; IPR001762; Disintegrin.

DR InterPro; IPR002870; Pep_M12B_propep.

DR InterPro; IPR001590; Reprolysin.

DR InterPro; IPR000884; TSP1.

DR InterPro; IPR000130; Zn_MTpeptdse.

DR Pfam; PF000090; tsp_1; 4.

DR Pfam; PF01421; Reprolysin; 1.

DR Pfam; PF01562; Pep_M12B_propep; 1.

DR SMART; SM00209; TSP1; 4.

DR PROSITE; PS00215; ADAM_MEROPS; 1.

DR PROSITE; PS00092; TSP1; 1.

DR PROSITE; PS00142; ZINC_PROTEINASE; FALSE_NEG.

DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.

[illegible]

[illegible]


```
Db 340 LLTRKDLCAAMNRCETGLSHVAGMCQPHRSCSINEDTGLPLAFTVAHELGHSGFIQHD 399
QY 373 -NVKVCVEFGKLRANHMSPITLIQIDRANPWSACSAIITDFLDGSHGDCILDDQSKP- 430
Db 400 GSGNDCEPV-GK--RPFMSQQLLYDAPLWTWSCRQYITRFLRGWGLCLDDPPAKDI 456
QY 431 ISLPEDLPASYSYLSQCELAFGVSGKPCPYM-QYCTKLMCTGKARGQVMVCOTRHFPWAD 489
Db 457 IDFPSVPGVLYDVSHQCRLOQYGAISAFCEMDMNVCHTLWCSVGT---TCHSKLDAADV 512
QY 490 GTSCEGKCLKACACVYHNLKRVDSWAKWDPTGPGSCGCGVQIARROCTNPTFA 549
Db 513 GTRGKKNWCLSGECPV-VGFRPEAVDGGWGSANWISCSRSCGMVGQSAERCTQPTPK 571
QY 550 NGKCYCEGVYKVRSCNLEPCPSASGKSPREOCQAFNHNSTNRLTLAVAWVPKYS 609
Db 572 YGRYCVGERKFRCLNQLACP--AGRPSEHVQCSEHFDAMLYKQGLHT----WVPVND 625
QY 610 VSPDRCKLITCRANGTYFVVLAPKVVVDGTLCSF--DSTSVCVQGVKICAKGCDNGLSKK 667
Db 626 VNP---CELCHRPANEFYFAKFLDRAVDVGTGTCYQVARSRLDCLINGICKNVGCDFTDSGA 682
QY 668 RFKCGVCGDKNCKKVTGLTKPMH-GYNFVVAIPAGASSIDIRQGYKGLIGDNYL 726
Db 683 MEDRCGVCHNGSTCHTVSTFEAEAGLVGVVDGLIPAGAREIRIQE-----VAEAAFL 737
QY 727 ALKNSQ-GKYLNGHFVVSVERDLVVKGLSLRYSGTGTAVERSLQASRLPLEPLTVEVL- 784
Db 738 ALRSEDEKPYFLNGWTTIQ-WNGDYQVAGTFTFYARRGN-WENLTSPGTPKEPVMQVPA 795
QY 785 -----SVGKMTPPRVKVSFYLPKPREDKSHDPDGPSPVHNSVLSNLSNVEQPD 837
Db 796 SRPGGSGRGVPRPSTLHGRSPRGVSPGSETEPGSEPCPAAASTSVS----- 845
QY 838 RPPARW---VA-----GSGWPCASCSGLQKRAVDGRGS-----AGQRTVPACDAARHP 884
Db 846 -PSLKPWNLVAHVHGWG--QAPLGLGWRHRLVLMGRLPTQLLFQESNGVHYEY-T 901
QY 885 VETQAGE-----PCTWELSAWSPCKSCGRG--FORSLKCVG 922
Db 902 IHREAGGHDEVPVPVPSWYHGWTKCTVTCGRGKGRHSPTCRG 946

RESULT 15
AT52_BOVIN STANDARD; PRT; 1205 AA.
AC F79331;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (pNPI).
GN ADAMTS2 OR NPI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=97225960; PubMed=9122202;
RA Collige A., Li S.W., Sieron A.L., Nusgens B.V., Prockop D.J.,
RA Lapiere C.M.;
RT "cDNA cloning and expression of bovine procollagen I N-proteinase: a
RT new member of the superfamily of zinc-metalloproteinases with binding
RT sites for cells and other matrix components.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).
RN [2]
RP PARTIAL SEQUENCE.
```

```
RX MEDLINE=95348096; PubMed=7622483;
RA Collige A., Beschin A., Samyn B., Goebels Y., Van Beeumen J.,
RA Nusgens B.V., Lapiere C.M.;
RT "Characterization and partial amino acid sequencing of a 107-kDa
RT procollagen I N-proteinase purified by affinity chromatography on
RT immobilized type XIV collagen.";
RL J. Biol. Chem. 270:16724-16730(1995).
CC -|- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
CC COLLAGEN BIOSYNTHESIS.
CC -|- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
CC alpha-1(I) at Pro-Gln and of alpha-1(II) and alpha-2(I) chains
CC at Ala-Gln.
CC -|- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
CC COLLAGEN TYPE XIV.
CC -|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -|- TISSUE SPECIFICITY: ENZYMACTIC ACTIVITY IS DETECTED AT HIGH LEVEL
CC IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS
CC AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS
CC WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -|- PTM: THE N-TERMINUS IS BLOCKED.
CC -|- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -|- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF DERMATOSPARAXIS, A
CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN
CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN
CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -|- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X96389; CAA65253.1; -
MEROPS; M12.301; -
InterPro; IPR001762; Disintegrin.
InterPro; IPR002870; Pep_M12B_propep.
InterPro; IPR001590; Reprolysin.
InterPro; IPR000884; TSP1.
InterPro; IPR000130; Zn_MTPeptidse.
Pfam; PF00090; tsp.1; 4.
Pfam; PF01421; Reprolysin; 1.
Pfam; PF01562; Pep_M12B_propep; 1.
SMART; SM00209; TSP1; 4.
PROSITE; PS00215; ADAM_MEPRO; 1.
PROSITE; PS00092; TSP1; 1.
PROSITE; PS00142; ZINC_PROTEASE; 1.
PROSITE; PS00427; DISINTEGRIN.1; FALSE_NEG.
Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.
SIGNAL 1 28 POTENTIAL.
FT PROPEP 29 253 BY SIMILARITY.
FT CHAIN 254 1205 ADAMTS-2.
FT METAL 402 402 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 403 403 BY SIMILARITY.
FT METAL 406 406 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 474 554 DISINTEGRIN-LIKE.
FT DOMAIN 555 611 TSP TYPE-1 1.
FT DOMAIN 612 715 CYS-RICH.
FT DOMAIN 717 845 SPACER.
```



```
FT DOMAIN 846 905 TSP TYPE-1 2.
FT DOMAIN 906 968 TSP TYPE-1 3.
FT DOMAIN 969 1024 TSP TYPE-1 4.
FT SITE 685 687 CELL ATTACHMENT SITE (POTENTIAL).
FT DOMAIN 31 35 POLY-ALA.
FT DOMAIN 177 180 POLY-GLU.
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 943 943 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 987 987 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1025 1025 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1092 1092 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1144 1144 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1205 AA; 133887 MW; 7B5B232A45320371 CRC64;

Query Match 23.18; Score 1192; DB 1; Length 1205;
Best Local Similarity 31.76; Pred. No. 2.4e-76;
Matches 329; Conservative 134; Mismatches 403; Indels 172; Gaps 42;

QY 1 MLLLGILTLAFAGTAG-----GFEPEREVVPRLD----- 32
DB 16 LLLLLLLPADARLAAAADPPGPGHGAEIRILAVEVRTDAQRLVSHVSAATAPAG 75
QY 33 -----PDINGRYYWRGPDSDQGLIFQITAFQEDFYHLTPDAQFLAPAFST 81
DB 76 VTRRAAPAIPLSG-----GSEEDPGRLFYNTVFGRLHLRLRNARLVAGATV 129
QY 82 EHLGVLQGLTGSSDLRCFYSGDV-NAEPDFAAVSLCGGLRGAFGYRGAEYVISP 140
DB 130 EQWGE--SGATRVPLLTGLVGVAGLAESSVALSNCDSLGLAGLRMBEERFIE 187
QY 141 NASAPAAQRNSQG-AHLQRGVP-----GGPSGDTSRGCVASGWNPAILLRAL 193
DB 188 KGL--AAKEAEGRVHYVHRTTSRPPPLGPGQALDT---GISADSLDSLAL-- 237
QY 194 RRAFGESRRRSRAKRFVSIPIRY-VETLVVADESVMKPHGAD-LEHYLLTLATAA 251
DB 238 --GVLEERVNSRRMRHAADDYNIIEVLGVDDSVQPHGTEHVQKYLTLNMVNE 294
QY 252 LYRHP SILNPINIVVVKVLLLRDSGPKV-TGNAULTLRFCAWOKKLNKVDKHP 310
DB 295 IYHDESLGAHINVLIRIILSYGKMSLIETGNPSOLENVCRWAYLQORQPTDHD 354
QY 311 DTAIFTRDLCGATCTDLGMADYGTMCDPKRSVIEDGLPSAFTTAHELGHVFNMP 370
DB 355 DHAIFLTRQDF-GPSGMQ--GVAPVTGMCHPVRSCTLNHEDGFSSAFVAHETG 411
QY 371 HDNV-KVCEEVFKLRANHMSPTLIQIDRANPWSACSAIITDFLDGHDCLLDQ 427
DB 412 HDGQGNRCGD---EVLRLGSIMAPLVQAAAFHFRHWSRCSQBELSYLHS--Y 466
QY 428 -SKPISLPEDLPASYTLSQCELAFAFGSKPCPTMOY---CTKLWCTGKAKGM 483
DB 467 HDWP-ALPO-LPGLHYSMNEQCRDFGLGYMMCTAFRTFPCQLWCS-HPDN 523
QY 484 HEPWADGTS CGGKCLKACV-----ERHNLKRVDSWAKWDPYGPCSRTC 536
DB 524 KGPDLGTWCAPGKHCFCGHCITLWLPDILKR-----DGNWGAWSPPGSC 575
QY 537 QLARRQCTNPTPANGKCYEGYRVKYSRNLPCPSSASGSKSFREEQCAFNGY- 590
DB 576 KPRTRQCDNPHANGRTCSGLAYDFQLCNSQDCPDALA--DFREEQCRQWDL 633
QY 591 -NHSTNRLTAVANVPKYSVSPRCKLICRANGTYFYVYAPKVVVDGILCS-P 648
DB 634 QHH-----WLP-HEHRDAKERCHLYCESKETGEVWSMKRWVHDGTRCS 682
QY 649 CVQGKCIKAGDNLGSKKRFDFCGVCGDNKSKKVTGLFTKPMH--GYNFVVAI 706
DB 683 CVRGDCRKVCGDVGIVSSKQEDKCGVCGDNDNCKVVKVGTFSRPKKLGILK 742
```

```
QY 707 SSIDIRQRYKGLIGDDNYLAKNSO-GKYLNGHFVYSAVERDLVYVKGSLLYSGTGA 765
DB 743 RHLIQEAD-----TSHHLAVKNLETKGKFLINEENDVPNSKTFITAMGVWEYRDE-DG 796
QY 766 VESLQASRPILPLTVEVLSYGMKMTPPRVRYSFYLPKPEPREDKSSHHPKDPGRFVLIHNSV 825
DB 797 RETLQTMGPLHGTITVLVIPEG---DARISLTY-----KYMIHEDS 834
QY 826 LSLSNQVQPPDRPPARWVAGSWGPGSCSGSLQKRAVDCRGSGAGORTYPA--CDA 883
DB 835 LNVDDNNVLEDDSVGYEAWALKKWSPCSKPCGGGQFTKYGCRRLDHKMVRHRCFCD 894
QY 884 P-----VETQACGEPCPTWELSAWSPCSKSGR-GFQRRSLKCV-----GHG 932
DB 895 PKAIRRTCNPCQECQ--PVWVTGEWPCSRSGRTGMQVRSVRCVQPLHNNTTTSVHTKH 952
QY 933 CNLHKBKPOELDFCVLRPC 950
DB 953 CN-DARPERRACNREL 969
```

Search completed: April 29, 2003, 17:13:55
Job time : 25.7002 secs

Db 89 GETLLLELLEQDGGVQVEGLAVOYLQGAPE--LLUGAEP--GTYLTGTINGDPESVASLHWD 145
Qy 121 GG-LRGAFGYGAEEYVISPPLNAPAAQNSOGAHLQRRGVPGGPGDPTSCRGVAGS 179
Db 146 GGALLGVQYRGAELHQLPGLGGTPNSA--GGPGAHLRRK-----SPASGGPGMCNV-- 196
Qy 180 WNPAILRALDPYKPRRAGFGESRRRSRGRKRFVSPRIYVETLVVADESVMVPHGADLE 239
Db 197 -----KAPLGSPSPRR--RAKRFASLSRFVETLVVADKNAAFHAGL 239
Qy 240 HYLLTLLTAARLYRHPHSILNPINIVVKKVLLLRDRSDGPKVTGNAALTIRNFCAMQKIL 299
Db 240 RYLLTVMAAAKARKHPSIRNPVSLVTVRLVILGSGEESPQVGPSAAQTLSRFCAMQRL 299
Qy 300 NKYSKHPFYWDTALLFTRQDLGATTCDTLGMADVTGMDCKPSCSVIEDDGLPSAFTT 359
Db 300 NTPEDSDPDHFTALFTQDLQCVSTCDTLGMADVTVCDPAPSCAIVEDDGLQSAFTA 359
Qy 360 AHELGHVFNMPHDNVKVCVEEYFGKL-RANIMHSPPTLIQIDRANPWSAAAITDFLDSG 418
Db 360 AHELGHVFNMLHDNSKPCISLNGPLSTRHVMAPVMAHVDPEPWSPCSAARFTDFLDNG 419
Qy 419 HGDCLLDPSKPSILPEDLPGASYTLSSQCELAFGVSGKPPYM-QYCTKLWCTGKAKGQ 477
Db 420 YGHLCLDKPEAPLHLVTFVTFPKQIDADRCQUTFGDPSRHCQPLPPCAALWCSGHLNGH 479
Qy 478 MYCQTRHPFWADGTSCEGKLCCLKGACVERHNLNKHVD--GSAKMDPYGPCSRTCQGG 535
Db 480 AMCQTKHSPWADGTPCGPAQACMGGRCLHMDQLQDFNPQAGMGWPMGMDCSRCTCGG 539
Qy 536 VOLAROCNTPTPANGGKYCEGVVYKRSCLNBPCCPSASGSKSPREEOCAFNGYHNSTN 595
Db 540 VQESSDCTRPVPRNGKYCEGRTRFRSCNTEDCP-IGSALTIFREEQCA---YNHRTD 595
Qy 596 RUTL---AVAWPYKGVSPDRCKLICRANGTGFYVLAIPKVDGTLCPSPDTSVCVQG 652
Db 596 LKSFEPGMDWYPRYTVGAPQDQCKTQCARALGYLYVLEPRVDGTPCSPDSSVCVQG 655
Qy 653 KCTKAGCDNGLSKRRFDKCGVCGDNKSKCKVTGLFTKPHGHYFNFFVATIPAGASSIDIR 712
Db 656 RCIHAGCDRIIGSKKFKDKMVCVGGDSCGSKSGSFRKFRGYGNVNVVTPAGATHILVR 715
Qy 713 QRCYKGLIGDDNYLALKNSQKYLNLNGHFVSVNAVERDLVVGSL-LRYSGTGTAVESLQA 771
Db 716 QQGNPG--HRSYTLALKLPDGSVALNGETLMPSTDVVLFGAVSLRYSGATAASETLSG 773
Qy 772 SRPLEPTVEVLSVGKMTTPRVRYSFLPKPREDKSHPKD 814
Db 774 HGPLAQTLQVLVAGNPQDTRLRYSEFFVPRPTPTPTPD 816

RESULT 3
T47158
hypothetical protein DKFzp762C11.0.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47158
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224379
A:Accession: T47158
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-550 <AAA>
A:Cross-references: EMBL:AL162080
A:Experimental source: adult melanoma (Mewo cell line); clone DKFzp762C11.0
A:Note: DKFzp762C11.0.1

Query Match 31.1%; Score 1604; DB 2: Length 550;
Best Local Similarity 49.7%; Pred. No. 1.2e-101;
Matches 284; Conservative 106; Mismatches 142; Indels 40; Gaps

A:Gene: CESP:F25H8.3
 A:Map position: 4
 A:introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 814/1; 854/1; 884/1; 914/1; 944/1; 974/1; 1004/1; 1034/1; 1064/1; 1094/1; 1124/1; 1154/1; 1184/1; 1214/1; 1244/1; 1274/1; 1304/1; 1334/1; 1364/1; 1394/1; 1424/1; 1454/1; 1484/1; 1514/1; 1544/1; 1574/1; 1604/1; 1634/1; 1664/1; 1694/1; 1724/1; 1754/1; 1784/1; 1814/1; 1844/1; 1874/1; 1904/1; 1934/1; 1964/1; 1994/1; 2024/1; 2054/1; 2084/1; 2114/1; 2144/1; 2174/1; 2204/1; 2234/1; 2264/1; 2294/1; 2324/1; 2354/1; 2384/1; 2414/1; 2444/1; 2474/1; 2504/1; 2534/1; 2564/1; 2594/1; 2624/1; 2654/1; 2684/1; 2714/1; 2744/1; 2774/1; 2804/1; 2834/1; 2864/1; 2894/1; 2924/1; 2954/1; 2984/1; 3014/1; 3044/1; 3074/1; 3104/1; 3134/1; 3164/1; 3194/1; 3224/1; 3254/1; 3284/1; 3314/1; 3344/1; 3374/1; 3404/1; 3434/1; 3464/1; 3494/1; 3524/1; 3554/1; 3584/1; 3614/1; 3644/1; 3674/1; 3704/1; 3734/1; 3764/1; 3794/1; 3824/1; 3854/1; 3884/1; 3914/1; 3944/1; 3974/1; 4004/1; 4034/1; 4064/1; 4094/1; 4124/1; 4154/1; 4184/1; 4214/1; 4244/1; 4274/1; 4304/1; 4334/1; 4364/1; 4394/1; 4424/1; 4454/1; 4484/1; 4514/1; 4544/1; 4574/1; 4604/1; 4634/1; 4664/1; 4694/1; 4724/1; 4754/1; 4784/1; 4814/1; 4844/1; 4874/1; 4904/1; 4934/1; 4964/1; 4994/1; 5024/1; 5054/1; 5084/1; 5114/1; 5144/1; 5174/1; 5204/1; 5234/1; 5264/1; 5294/1; 5324/1; 5354/1; 5384/1; 5414/1; 5444/1; 5474/1; 5504/1; 5534/1; 5564/1; 5594/1; 5624/1; 5654/1; 5684/1; 5714/1; 5744/1; 5774/1; 5804/1; 5834/1; 5864/1; 5894/1; 5924/1; 5954/1; 5984/1; 6014/1; 6044/1; 6074/1; 6104/1; 6134/1; 6164/1; 6194/1; 6224/1; 6254/1; 6284/1; 6314/1; 6344/1; 6374/1; 6404/1; 6434/1; 6464/1; 6494/1; 6524/1; 6554/1; 6584/1; 6614/1; 6644/1; 6674/1; 6704/1; 6734/1; 6764/1; 6794/1; 6824/1; 6854/1; 6884/1; 6914/1; 6944/1; 6974/1; 7004/1; 7034/1; 7064/1; 7094/1; 7124/1; 7154/1; 7184/1; 7214/1; 7244/1; 7274/1; 7304/1; 7334/1; 7364/1; 7394/1; 7424/1; 7454/1; 7484/1; 7514/1; 7544/1; 7574/1; 7604/1; 7634/1; 7664/1; 7694/1; 7724/1; 7754/1; 7784/1; 7814/1; 7844/1; 7874/1; 7904/1; 7934/1; 7964/1; 7994/1; 8024/1; 8054/1; 8084/1; 8114/1; 8144/1; 8174/1; 8204/1; 8234/1; 8264/1; 8294/1; 8324/1; 8354/1; 8384/1; 8414/1; 8444/1; 8474/1; 8504/1; 8534/1; 8564/1; 8594/1; 8624/1; 8654/1; 8684/1; 8714/1; 8744/1; 8774/1; 8804/1; 8834/1; 8864/1; 8894/1; 8924/1; 8954/1; 8984/1; 9014/1; 9044/1; 9074/1; 9104/1; 9134/1; 9164/1; 9194/1; 9224/1; 9254/1; 9284/1; 9314/1; 9344/1; 9374/1; 9404/1; 9434/1; 9464/1; 9494/1; 9524/1; 9554/1; 9584/1; 9614/1; 9644/1; 9674/1; 9704/1; 9734/1; 9764/1; 9794/1; 9824/1; 9854/1; 9884/1; 9914/1; 9944/1; 9974/1; 10004/1; 10034/1; 10064/1; 10094/1; 10124/1; 10154/1; 10184/1; 10214/1; 10244/1; 10274/1; 10304/1; 10334/1; 10364/1; 10394/1; 10424/1; 10454/1; 10484/1; 10514/1; 10544/1; 10574/1; 10604/1; 10634/1; 10664/1; 10694/1; 10724/1; 10754/1; 10784/1; 10814/1; 10844/1; 10874/1; 10904/1; 10934/1; 10964/1; 10994/1; 11024/1; 11054/1; 11084/1; 11114/1; 11144/1; 11174/1; 11204/1; 11234/1; 11264/1; 11294/1; 11324/1; 11354/1; 11384/1; 11414/1; 11444/1; 11474/1; 11504/1; 11534/1; 11564/1; 11594/1; 11624/1; 11654/1; 11684/1; 11714/1; 11744/1; 11774/1; 11804/1; 11834/1; 11864/1; 11894/1; 11924/1; 11954/1; 11984/1; 12014/1; 12044/1; 12074/1; 12104/1; 12134/1; 12164/1; 12194/1; 12224/1; 12254/1; 12284/1; 12314/1; 12344/1; 12374/1; 12404/1; 12434/1; 12464/1; 12494/1; 12524/1; 12554/1; 12584/1; 12614/1; 12644/1; 12674/1; 12704/1; 12734/1; 12764/1; 12794/1; 12824/1; 12854/1; 12884/1; 12914/1; 12944/1; 12974/1; 13004/1; 13034/1; 13064/1; 13094/1; 13124/1; 13154/1; 13184/1; 13214/1; 13244/1; 13274/1; 13304/1; 13334/1; 13364/1; 13394/1; 13424/1; 13454/1; 13484/1; 13514/1; 13544/1; 13574/1; 13604/1; 13634/1; 13664/1; 13694/1; 13724/1; 13754/1; 13784/1; 13814/1; 13844/1; 13874/1; 13904/1; 13934/1; 13964/1; 13994/1; 14024/1; 14054/1; 14084/1; 14114/1; 14144/1; 14174/1; 14204/1; 14234/1; 14264/1; 14294/1; 14324/1; 14354/1; 14384/1; 14414/1; 14444/1; 14474/1; 14504/1; 14534/1; 14564/1; 14594/1; 14624/1; 14654/1; 14684/1; 14714/1; 14744/1; 14774/1; 14804/1; 14834/1; 14864/1; 14894/1; 14924/1; 14954/1; 14984/1; 15014/1; 15044/1; 15074/1; 15104/1; 15

N:Alternate names: procollagen N-proteinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

Matches	347;	Conservative	128;	Mismatches	338;	Indels	132;	Gaps	24;	
QY	101	CFYSGDVNAEPD	SFAA	SILC---	GGLRGAFYRGA	EYVISLP	NASA---	PAAORNSQGA	154	
Db	169	CYIYRAHVKG	-VQHHSIVN	CLDSEDGL	YGLMAL	PSCIHTH	VEPIISGNGTEH	DGASRHRQ---	225	
QY	155	HLLQRRVGP	GGSGDPT	RCGY-----	ASGNPAILRAL	DPYKPRRAG	FGESRSR	204		
Db	226	HLVRKFDPMH	KFSDFHLN	STSVNETT	TVATWQD	WEDVIER-----	KARS	272		
QY	205	RRSGRAKRFVS	IPRYVETLV	VVADESVM	KFGADLEHY	LLTLATAARLY	RHPSTLNP	INI 264		
Db	273	RAAN-----	SWDHVVEVL	VADTKMYEYH	GRSLEDIV	LTFLSFVASIYR	HQSURASIN	326		
QY	265	VYVKVLLLRD	SRDSPKVT	GNAAALTR	NFCAMQK	KLNVSKDHP	EYWTAILFT	RQDLCGA 324		
Db	327	VYVKLIVL	TENAGMP	RIITONAAQ	TLQDFCRWQ	QYVNDPDDSSVQ	HHDAVILL	TRKDICRS 386		
QY	325	T-TCDTGLM	ADVGNCDP	KRSCSVIED	GLPSATTAH	ELGHVNM	PHDVKVCEV	--- 380		
Db	387	QKCDTGLA	ELGTMDCK	SAIEDNGL	SAAFTIAH	ELGHVFSI	PHDERKCS	YMPV 446		
QY	381	-----FG	KLAN---	HMSPTLIQ	IDRANPWS	SACSAAILT	DFLDSGHG	-DCLLD 425		
Db	447	NKVCFQST	KEDKTKQ	FQNNFHIM	APTLEYNTH	PSWSPCSAG	MURLENNR	GOTCQLFD 506		
QY	426	QSPKPSIL	PE---	DLPGASYTL	SQOCBELAF	GVSGKPCPYM	QYQYCTKL	WCTGKAKGMVQC 481		
Db	507	QPVERRYED	VFRDEPG	KYKDAHQ	QCKFVFG	PASELCPYMT	CRRLWCAT	FYGSQMGR 566		
QY	482	TRHPFWAD	GTS	CGRGK--L	CLKGACVE	RHNLUNKH	RVDSWAK	WDYPGCSRT	CGGVOLA 539	
Db	567	TOHPFWAD	GTPCDES	RSMFCHG	ACVRLAP	ESLTKIDQW	GDWRSWG	ECSTRCTGGVQKG 626		
QY	540	RRQCTNPT	PANGKYCE	GVRYKSC	NLEPCPSS	ASGKSPRE	EOCEAFNGYN	-----HS 593		
Db	627	LRDCSPK	PRNGKYC	VGOREYR	SCNTQEC	PWDT--QPYRE	QVCSFENK	NDIGIQGVAS 684		
QY	594	TNRULTAV	AWVPKYS	GVSPRDK	CKLICRANG	TGYFYVLAP	KVVDGTL	CLSPDSTSV	CVQKG 653	
Db	685	TN-----	THWPKYAN	VPNERAP	NERKLYC	RLSGSAAFY	LLRDVKV	VDGTPCDR	GGDDICVAGA 739	
QY	654	CIAKACDGN	LGSKRFR	DKCGVGD	NKSCKVYTG	LTKP-MHGN	VNFVAIP	PACASSIDIR 712		
Db	740	CMPAGCDH	LHSTLRR	DKCGVGD	SDSCKVVGK	TFFNQGT	FYGYNEVM	KYMPAC	SANIDIR 799	
QY	713	QRGYKGLT	GDNYLL	ALKN	SO	KYLLNGH	FVVSAVERDL	VKLSLLRY	SGTGTAVESLQAS 772	
Db	800	QKGYNNK	EDDNYL	SRAANG	EFLNGH	FQVSLARQ	IAFQDVI	LEVYSGS	DAIERINGT 859	
QY	773	RPILPLET	VEVLSV	GKMT	PPRYSYF	YLKPRE	-----	-----	806	
Db	860	GPIRSDIY	VHVL	SVGS-HP	PDISEY	MTAAVP	NAVIRP	ISSALYL	WRVTDTWTECDRACR 918	
QY	807	-----DK	SHPKDPR	GPSVLH	---SVL	SLNSQVEQ	DDPRP	PABWAGS	WGPCS 852	
Db	919	GOQSKLM	CLDM	STHROS	-----	HDRCN	QNVLPK	QATRM	CNIDCS	TRWITVEDYSSCS 971
QY	853	ASCSGGLQ	KRAVDC	RGSAG	ORTVPA	---CDA	AHRPVET	QACG	PCP--TWEL	SAPSPCS 906
Db	972	AKCSGGK	RQKRV	SCVKW	EGDQTP	ASEHLCD	RNSKPS	DIASCY	IDCSGR	KWNGEWTSCS 1031
QY	907	KSCG-RG	FQRSLK	CVCHG	GRLLR	ADOCNL	LRKPQ	EOLF	FCVLRPC	950
Db	1032	ETCSGK	MHRKSY	CVYD	SDSNRRY	DESJLC	RGKEAT	ERECNR	ITPC	1076

RESULT 5

C:Accession: T18517
R:Collage, A.: Nussgens, B.V.; Lapiere, C.M.
submitted to the EMBL data Library, February 1996
A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.
A:Reference number: Z18941
A:Accession: T18517
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1205 <COL>
A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1
A:Experimental source: skin
C:Genetics:
A:Gene: PC I-NP
C:Function:
A:Description: catalyzes cleavage of the propeptides of type I and II collagens prior to
C:keywords: hydrolase; metalloproteinase

Query Match 23.1%; Score 1192; DB 2; Length 1205;
Best Local Similarity 31.7%; Pred. No. 3.9e-73;
Matches 329; Conservative 134; Mismatches 403; Indels 172; Gaps 42;

Qy 1 MLLGLILTLAFAGRTAG-----GFEPEREVVVPIRLD----- 32
Db LLLLLLLPADARLAAAADPPGGQGHGAERILAVPVRTDAQRLSHVSHVSAATAPAG 75
Qy 33 -----PDINGRRYYWRGPDSDGGLIFQIFATAQEDFYHLHTPDQAFLAPAFST 81
Db VTRRAAPAQITPLSG-----GSEEDPGGRLEFYNVTGFRDLHLRLPNARLAVAPCATV 129
Qy 82 EHLGVPLQGLTGSSDLRRCFYSDV--NAEPDSFAVSLGGLRGAGFYGAEBYVLSPL 140
Db 130 EMOGE--SGATRVPEPLGTCLYVDVAGLAESSVALSNCGLAGLRMEEEFFIEPLE 187
Qy 141 NASAPAAQRNSQG--AHLQRRGVP-----GPGSDPTSRGVSAGWNPAILRALDPYXP 193
Db 188 KGL--AAKEAPQGRVHVYHRTTSRPPPLGGPQALDT---GISADSLDSLRAL----- 237
Qy 194 RRAQGESRRRSRRGRKRFVSPTRY--VETILVADSWAFHGAD--LEHVLILLATAAR 251
Db 238 ---GVLEERVNSSRRRRHAADDYDIEVLLGVDDSVQFQGHTEHVQKYLTLMLNIVNE 294
Qy 252 LYRHSILNPINIVVVLRLDRSDGPKV--TGNAAITLNFCAWQKLNKVDKHPPEY 310
Db 295 IYHESLGAHNVVLVRIILSYKSLSLTEIGNPSQSLNVCWVAQVLOKPDTHDEYH 354
Qy 311 DTALFTRQDLGGATTCDTLGMADVTGMDPKRSCSVIEDGLPSATTAHGLHGVNMP 370
Db 355 DHAIFLTRQDF--GPSGMQ--GYADVTGMCHPVRSCITLNHEDGFSSAFVAHETGHLGME 411
Qy 371 HDNV--KYCEEVFGKLRANHMSPTLIQIDRANPHSACSAALITDFLDSGHGDCLLDOP-- 427
Db 412 HDGQNRGCD---EVRLGSIAPLVQAFAHFRHWSRCSQELSRYLHS--YDCLURDDPFT 466
Qy 428 -SKTSLPEDLPGASYTLISOQELAFVGSGKPCPYMY---CTKLWCTGKAGKQWQOTR 483
Db 467 HDWP--ALPQ--LPLGLHYSNNEQCFDGLGYMCTAFRTFDPCKQLWCS--HPDNPYFCKTK 523
Qy 484 HFPNADGTCGEGKLCILKAGCV-----ERHNLNKHVRGWSWAKWDPYGPCSRCTCGGV 536
Db 524 KGPPLDGTMCAPGKHCFKGHCILWLTPLDKR-----DGNMGAWSPFGSCSRCTCGTV 575
Qy 537 QLARQCTNPPTPANGKKYCEGVRYKRYSCNLEPCPSSASGSKSFREEOCEAFNGY----- 590
Db 576 KFRTRQCDNPANGRTCSGLAYDFQLCNSQDCPDALA--DFREOCROWDLYFEGHDA 633
Qy 591 -NHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYFYVLAPKVDGTLCS--PDSTSV 648
Db 634 QHH-----WLP--HEHRDAKERCHLYCESKETGEVVSQKRMVHDGTRCSYKDAFSL 682
Qy 649 CVQGCKIKAGCDNGLSKKRFDKGCVCGDGNKSKKVTGLFTKPMH--GYNFVVAIPAGA 706
Db 683 CVRGDCRVGCDVIGSSKQEDKCGVCGDGNSHCKVAVKGTFSRSPKILGYIKMFEIPAGA 742

Qy 707 SSIDIRQGYKGLIGDDNYLALKNSQ--GKYLLNGHFVYSAVERDLVVKGSLLRYSGTGTA 765
Db 743 RHLTIQEAD-----TSHLLAVKNELTGKFLINEENDVDPSKTFIAMGVWEYERDE-DG 796
Qy 766 VESLQASRPILTEVLTVSVGKMTPPRVRYSFYLPKPREDKSSHPKDPGRGSPVLHNSV 825
Db 797 RETLQTMGPLHGTITVLVIPEG---DARISLTY-----KYMIIHEDS 834
Qy 826 LSLSNQVQPDPRPPARVWAGSWGSPCASCSGSLQKRAVDCRGSGAGQRTVPA--CDAHR 883
Db 835 LNVDDNNVLEDSVGYEWALKKWSKPCPKCGGSGQFTKYGCCRRLDHKMHVRGFCDSVSK 894
Qy 884 P-----VETQACGEPCTWELSAWSPCSKSGR--GFORRLKCV---GHGGRLARDQ 932
Db 895 PRAIRTCNPOECQ--PWWTGWEPSCSGRTGMQVRSVRCVQPLHNTTTSVHTKH 952
Qy 933 CNLHRKPOELDFCVLRPC 950
Db 953 CN-DARPEGRACNRELC 969

RESULT 6
C89114
protein C37C3.6a [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: C89114
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: C89114
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1558 <STO>
A:Cross-references: GB:chr_V; PIDN:AAC25857.1; PID:g3294501; GSPDB:GN00023; CBSP:C37C
C:Genetics:
A:Gene: C37C3.6a
A:Map position: 5

Query Match 13.2%; Score 681; DB 2; Length 1558;
Best Local Similarity 33.7%; Pred. No. 3.8e-38;
Matches 162; Conservative 63; Mismatches 176; Indels 80; Gaps 19;

Qy 517 GSAKWDPPGCSRCGGGVQVQLARQCTNPANGKYCEGVRYKRYSCNLEPCPSSASG 576
Db 77 GNMGPVWPENECSRCGGGVQLEKQCSGD-----CTGASVRIISCNLNAC---ESG 125
Qy 577 KSFREQCEAFN-----GYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYFYVL 631
Db 126 TDFRACQCKFNDALDGNVHK-----WTP-YKG---KNKELCKPESGNFYKWK 172
Qy 632 APKVDGTLCSPTDSVVCVQKCIKAGCDNGLSKKRFDKGCVCGDGNKSKKVTGLFTK 691
Db 173 ADKVDGTRKDSKSDICVDGECPLVGCGLGSSLFKDKCGCDGSGTCTIEGREDE 232
Qy 692 P--MHGYNFVAIPACASSIDIRQYKGLIGDDNYLALKNSQKYLNGHFWWSAVERD 749
Db 233 RNLSPGYHDIILPEGATNIKIQEAR-----KSTNNLALKNGSDHYFLNGSLIO-VEKE 286
Qy 750 LVVKGSLRYSGTGTAVESLQASRPILTEVLTVSVGKMTPPRVRYSFYLPKPREDE-- 807
Db 287 VEVGGTIFVYD--DAEPETLSAQGPLSELTVALLPKRSRDTAIFYESIPLEEDVDM 344
Qy 808 -----KSSHPKDPGRGSPVLHNSVLSLSNQVEQDDRP-----PAR 842
Db 345 YKEDNTPCWSVSGKGVQTRNLYCIDGKNKGRVEDDLCEENNATKPEFEKSETVDCEAE 404
Qy 843 WVAGSGPCPSASCG--SLQKRAVDC--RGSAGQRTVPACD---AAHRPVETOACGE-PCPT 896
Db 405 WFTGDMESCSSTCGDQGGQYRVVYCHQVANGRRVTVEDGNCTVPRPPVQTCNRFACPE 464

QY 897 WELSAWPCSKGCGRFRRLK-----VHGGRLLARDQCNLHRKQELD---FCVLRP 949
 Db 465 WQAGPWSACSEKCGDAFYRSVTCRSEKEGEGKLLAADACPADQ- EKFDPTERTCNLGP 523

QY 950 C 950
 Db 524 C 524

RESULT 7

T18856
 angio genesis inhibitor homolog - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 01-Dec-2000
 C:Accession: T34395; T34394
 R:Geisel, C.; Bradshaw, H.
 submitted to the EMBL Data Library, July 1996
 A:Description: The sequence of C. elegans cosmid C37C3.
 A:Reference number: Z21518
 A:Accession: T34395
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2167 <GEI>
 A:Cross-references: EMBL:U64857; PIDN:AAC25868.1; GSPDB:GN00023; CESP:C37C3.6b
 A:Experimental source: strain Bristol N2; clone C37C3
 A:Accession: T34394
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1555, SRP' <G82>
 A:Cross-references: EMBL:U64857; PIDN:AAC25867.1; GSPDB:GN00023; CESP:C37C3.6a
 A:Experimental source: strain Bristol N2; clone C37C3
 C:Genetics:
 A:Gene: CESP:C37C3.6b; CESP:C37C3.6a
 A:Map position: 5
 A:Introns: 32/3; 104/2; 156/2; 207/1; 459/2; 536/3; 577/2; 1105/3; 1367/1; 1438/1; 1556/

Query Match 13.2%; Score 681; DB 2; Length 2167;
 Best Local Similarity 33.7%; Pred. No. 5.7e-38;
 Matches 162; Conservative 63; Mismatches 176; Indels 80; Gaps 19;
 QY 517 GSAWWDYPCSRFCGGVQLARQCNPTPANGKVCYGVRYKRCNLEPCPSSASG 576
 Db 77 GNGFWPVENECRSGGVQLEKQCSD-----CTGASRYIISCNLAC---ESG 125
 QY 577 KSFEEQCEAFN-----GYNHSTNRLTLAVAWPKYSGVSPRDKKLCIRANGTYEVL 631
 Db 126 TDFAEQCSKFDEALDGNHYK-----WTP-YKG---KNKELVCKPESGNFYKW 172
 QY 632 APKVVDGTLCPDSTSVQVQKCIKAGCDGNLGSKKRFDKCGVCGDNKSKYVTLGFTK 691
 Db 173 ADKVVDGTLCKDSKSNIDCVGECLPVGCDGKLGSLKPKDCKGCDGSGTCKTIEGRFDE 232
 QY 692 P---MHGYNFWVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQKYLNLNGHFVVSVERD 749
 Db 233 RNLSPGYHDIKLPGANIKIQR-----KSTNNLALKNGSDHFYLYNGGLIQ-VEKE 286
 QY 750 LVVKGSLRLYSGTGTAVESLQASRPILRLPVEVLVSGKMTPPRVRYSFYLPKPEPRD-- 807
 Db 287 VEVGCTIFVD--DAEPETLSAQGPLSEELIVALLFRKGRSDTALKYEFSTPLEEVDYM 344
 QY 808 -----KSHPKDPKPGPSVLHNSVLNSQVQEPDRP-----PAR 842
 Db 345 YKFDNWTPCSVSCGKGVQTRNLYCIDGKNKGRVEDDLCEENNAKPKPEKSCETVDCAE 404
 QY 843 WYAGSWGPCASGCG-SGLQKRAVDC-RGSAGRTVPACD-----AAHRPVETQACGE-PCPT 896
 Db 405 WFTGDWESCSTCGDGOQYRVVCHQVFANGRRVTVEDGNCVTERPVVKOTNRFPACPE 464
 QY 897 WELSAWPCSKGCGRFRRLK-----VHGGRLLARDQCNLHRKQELD---FCVLRP 949
 Db 465 WQAGPWSACSEKCGDAFYRSVTCRSEKEGEGKLLAADACPADQ- EKFDPTERTCNLGP 523

QY 950 C 950
 Db 524 C 524

RESULT 8

T18856
 angio genesis inhibitor homolog - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T18856; T24653
 R:McMurray, A.
 submitted to the EMBL Data Library, July 1995
 A:Reference number: Z19031
 A:Accession: T18856
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1444 <WIL>
 A:Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN00028; CESP:C02B4.1
 A:Experimental source: clone C02B4
 R:McMurray, A.
 submitted to the EMBL Data Library, July 1995
 A:Reference number: Z19917
 A:Accession: T24653
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1444 <WIL>
 A:Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:C02B4.1
 A:Experimental source: clone T07C5
 C:Genetics:
 A:Gene: CESP:C02B4.1
 A:Map position: X
 A:Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3;

Query Match 10.8%; Score 559.5; DB 2; Length 1444;
 Best Local Similarity 22.0%; Pred. No. 6.7e-30;
 Matches 263; Conservative 127; Mismatches 400; Indels 403; Gaps 52;

QY 43 RGPEDSGDGLITQITAFQEDFYHLTPDAQFLAP-AFSTEHLGVLPQGLTGSSDLRRC 101
 Db 36 RHPLKNGN--LKMKFTAMNDTYHLNLRKNSRIVSPHIIISVVRHGDDDDVTYAGLRDYEQC 93
 QY 102 FYSGDVNAEPDFAAIVSLCGLRGAFGYRGAEEVVISPLPNASAPAAQRNSOGAHLQRG 161
 Db 94 HYQGEVKNHGMKRAAISDCGALMGSIVMEDHFLVQLTPK----RVHHLQKHEHLVYKRS 149
 QY 162 VPGGSPGDPTRSGCVASGWNPAILRALDPYKPRRAGFGESRRSRGRKRFVSIPIRY-- 219
 Db 150 -----AGLLTNAESKIREETRLQEEQESFCDT-SEQLDDPA---MTIPAHLLH 193
 QY 220 ---VEFLVVADSESMYKFHGAD-----LEH-----YLLTLL 246
 Db 194 FNTIPTISAQLDSSF1-FPNMDDPITLIEIGLFLDSKLFHEHFEYIQDAEQHLLFSLALI 252
 QY 247 ATAARLYRHSIILNPINIVVVKVLLLRDRS----GPKVTGNAALTFLRNCAMQKLNKV 302
 Db 253 NNHVLYQQDTLPNLDIVIVRYEMWRTOPSAISTGVHKNGOAQSLDDAFRCVQAHHNPG 312
 QY 303 SD-KHPEYWDTAILFTRODLCGATTCTDGLMADVGTWCDPKRSCSVIEDDGLPSAFTTAH 361
 Db 313 TDTDMNHVDHGLLTGYDIYHTTT-SVAGVAPVARMCDPLFACSLVEGLHGRSEVLAH 371
 QY 362 ELGHVFNPMHDNVK-VCEEVFGKLRANHMSPTLIQRANPNASACSAALITDFL----D 416
 Db 372 EMHNNMGVHDGVQVQNCNCCCLMSAVNAGKT-----TWSDCSVREFNAFLQLDLE 423
 QY 417 SGHGDCILLQDQSPKPSILPE----DLPGASYTLSSQCELAFGVGSK-PCP-----YMOYCTK 467
 Db 424 SGRGNCLRDASPLGISTNHLSDLRLPQRTAQOCSYFWRDVKVEIPNGRAMDDICRI 483
 QY 468 LWCTGRAKGMQVQCTRHFFWADTSCGEGKCLKLKACVE-RHNLNKHVRVDGSGNAKWD--- 523
 Db 484 LWCNGSGS---TISTAH-PALEGSGWCGANKWCHKGQCTHTWTFGLTPVPIDGSEWSEGGAE 539

Qy 524 ---PYGCSRCTCGGVLARROCTNPTPANGKGYCEGVRYKYSRNCLEPCSSASGSKFR 580
 Db 540 KGCPQOQCAVSGSTTVGQHRDVCNPNNGKTCGEGANRIGVCG---ATSSNCLGFT 595
 Qy 581 EQO-----CBFA-----NGYNISTNRLTLAVAWPKYSGVPRDKCKLICRA 622
 Db 596 REEFGNKCISIKYHPKPDQQLTGEGFEHST-----OPCRVWCHL 636
 Qy 623 NGT-----GYFYVLPKVVVDGLCPDSTSVCGVCKIKAGCDGN----- 662
 Db 637 IGSELINKQGF-----PDGTPCGFD--AYCVGGOCIALSCDNKALVEQPCDRIEG 687
 Qy 663 ----- 662
 Db 688 RSVHWEWSSWSRCSVSCGLGGRVREKRSRCKOGVSESRCEGVLRDCEBFGW 747
 Qy 663 -----LSKKRF-----DKG-----VCGG-- 677
 Db 748 KEMGSCSEKALGVQKRFRCLTDOSKHLQBERPCDNEGCVTNWDEWSSCSQSCGGGR 807
 Qy 678 -----DNKSKCKVTGLFTKPMHGYNFVVAIPAGAS---SIDIRORGYK 717
 Db 808 RYRTRKLDKDDGDDLEKESCN-----TQKCSOSWGDWLPSCVSCGIGFQIRER--- 858
 Qy 718 GLIGDDNYLALKNSQGYKLLNGHFVWSAVERDLVWKGSLRYS----- 760
 Db 859 --LCDGELCATANKQAR--TCNQOQCPSAF-----SLSWSEWGEWTTCSATCGEGLO 907
 Qy 761 -----GTGTAVESLQASRPILPVTVEVLSVGMKTPPRVRYSYFLPK----- 803
 Db 908 SRERSCRRGSCDTEDDASQTRCNGVPCHESYLTWSEWTTCTCSEFSDSRKRIAKCDGTTE 967
 Qy 804 -----PREDKSSHPKDPGSPVLSHNSVLSNOVEOPDDRP-----P 840
 Db 968 NCQDKIDEETCDIACLEKHSIFGIPSRPKLI-----TSNLRKAFGRPLPIBSIHS 1021
 Qy 841 ARWAGSWGPCASCSGLQAKRAVDRCSAGQRTVPACDAARHPRVETQACG--EPCFTWEL 899
 Db 1022 EKW--SEWGPCSVTCGSGRRVTRGQCEA-----SCPEQH--IQTECNLNSC--LEL 1068
 Qy 900 ---SAWSPCKSCGR--GFQRLSKCVGHGGRLLARDQCNLH---RKPOELDFC 945
 Db 1069 FIWDSWSSCSKSGQDGIQTRQKLC-----LFNNAECSSYAESRRCKDLPSC 1115
 RESULT 9
 T16892
 hypothetical protein T19D2.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T16892
 R:Bentley, D.
 submitted to the EMBL Data Library, December 1995
 A:Description: The sequence of C. elegans cosmid T19D2.
 A:Reference number: Z18599
 A:Accession: T16892
 A:Status: preliminary;
 A:Molecule type: DNA
 A:Residues: 1-860 <BEN>
 A:Cross-references: EMBL:U42846; NID:g1125809; PID:g1125810; PIDN:AAA83600.1; CESP:T19D2
 C:Genetics:
 A:Gene: CESP:T19D2.1
 A:Introns: 56/3; 96/1; 171/3; 199/3; 245/2; 293/3; 335/3; 426/3; 484/2; 505/3; 58
 Query Match 9.9%; Score 510.5; DB 2; Length 860;
 Best Local Similarity 26.1%; Pred. No. 7.9e-27;
 Matches 198; Conservative 92; Mismatches 310; Indels 159; Gaps 40;
 Qy 220 VETLVVADESM-----VKFHGADLEHYLLTLATAAR-----LYRHPSTLNIPINIVVVKLL 271
 Db 135 VELAVFADAMWDHFKKMYKAAENMHTFTFMVNNIDVLTQRLQPRINKIVRYEI 194

Qy 272 LRD-----RDSGPKVTGNAALTRNFCAMQKLNKVKSDKHPYWDTAILFTQDLCGATTC 327
 Db 195 LKNIPHLNARKSHSGDYDRLLDAFCQYQNEINPNPDPRHWHALLFSYDILHRNGVK 254
 Qy 328 DFLGADYGTMCDDPKRSCSVTDDGLPSAFTTAHELGHVFNMPHDNVKVCSEVEFGKLRAN 387
 Db 255 TVAGYAPVKGMCVRSCTINEGLDFGVFVVTHEMGLSGLMYHDGNECD-----LRC- 308
 Qy 388 HNSPTLLIQIDRANPWSACSAAITDLDLSDGH-GD-----CLLD-----QPSKPLSLPE-D 436
 Db 309 CIMSPSV--GSGKTHWSQCSVNEAMATFV--GHLGDFFRPPNCLQDASANEQMAFKESE 364
 Qy 437 LPGASYTLISQOCELAFGVGSKPC-----PYMQYCTKLWCTGKAKGOMVCOCTRHPW 487
 Db 365 PPGQLFTLDEQCEIFHG-----ECMKHELDQOTMONICOMVWC-----GNGEVIRTAH-PA 416
 Qy 488 ADGTSCGEGKCLKGACVHERHNLKHRVDGSGWAKWD--PYGPCSRTCG-----GVQOLAR 540
 Db 417 LEGTYCGFMICRQOCQVSSQLMRVTV--GGWSTWDRPAPTCCGRCSCQCEIRGOIRMR 475
 Qy 541 --ROCTNPTPANGKGYCEGVRYKYSRNCLEPCSSASGSKSFREOCEAFNGYNNHSTNRLT 598
 Db 476 SIROCNPSSNNGGAPCOGDEARGMVCHRDVC--NGDSIENYATRVCSRLRDENAIPTI- 533
 Qy 599 LAVAWPKYSGVS---PRDKCKLICRANGTYFYVLAIPKVVVDGLTCLSPDSTSVCGQKCI 655
 Db 534 -----LSGEGMFEQAMCKIWLISGTSNIRTVS--NFPDGPAGCPGG--QYCIKIGECR 582
 Qy 656 KAGCDNGLSKRRFKCGVCGGDNCKKVTGLFTKPMHGYNFVVAIPAGASSIDIRORG 715
 Db 583 PLLCGSTILAYSEAD--CPL-----SVLQTT--TTPPHVHVS----- 617
 Qy 716 YKGLIGDDNYLALKNSQGYK-----LLN---GHFVVS--AVERDLVWKGSLRYSG--TGTA 765
 Db 618 -----DQFACKTNPYKEHKHTPFLNEWSGWSVSECVTYDCHTQGVKVRVRRCLAGVC 670
 Qy 766 VESLQASRPILPVLVEVLSVCKMTPPR--VRYSYFLKEPEDEKSSHPKDPGRGVSVLHN 823
 Db 671 AGALRERQCTPCPGSERPL--TTPSPQOQTFNRFTIAPLPNQI-----N 714
 Qy 824 SVLSLSNQVEQPDPPARWAGSWGPGSCASGSLQ--KRAVDRCGSAGQRTVPACDAAH 882
 Db 715 MIL-----RKVDHW--GPHSACSVTCGTGKLRRENCIGQCEAETGPCV---- 757
 Qy 883 RPVETQACGEPCTW--ELSAWSPCKSCGGRGFRSLKC 920
 Db 758 ---MQSCRENKNTWTESQSDCSVNCGEVQVFRKRC 792
 RESULT 10
 T00260
 hypothetical protein KIAA0605 - human
 C:Species: Homo sapiens (man)
 C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
 C:Accession: T00260
 R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara,
 DNA Res. 5, 31-39, 1998
 A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp
 A:Reference number: Z14086; MUID:98290545; PMID:9628581
 A:Accession: T00260
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-951 <NAG>
 A:Cross-references: EMBL:AB011177; NID:g3043733; PIDN:BAAX25531.1; PID:g3043734
 A:Experimental source: brain
 C:Genetics:
 A:Note: KIAA0605
 C:Superfamily: thrombospondin type 1 repeat homology
 F:46-106/Domain: thrombospondin type 1 repeat homology <THRL>
 Query Match 9.5%; Score 493; DB 2; Length 951;
 Best Local Similarity 22.2%; Pred. No. 1.4e-25;
 Matches 171; Conservative 55; Mismatches 183; Indels 360; Gaps 27;

QY 517 GSWAKWDYPGCSRTCCGGVQLARQC---TNPTPANGKYCEGVRYKYSCHNLEPCPS 572
DQ 51 GWTWTAF---SRCCGGVTSQERHCLQORRKSVPGNRTCTGTGSKRYQLCRVQEC- 106
QY 573 SASGKSFRECEAFNG---YNHSTNRLTLAVAWPKYSGVSPRD-----KCLICRAN 623
DQ 107 -PDGRSFRECEQVSNHYNRTHQ-----WKPLY-----PDDVHLSRFPCLHC-TT 154
QY 624 GTGYFYVLAPKVVDTGLCS-PDSTSVCVQGCIRAGCDNGLGSKRRFDCGVCGGDNKSC 682
DQ 155 VDGQQLMVP-AROGTSKLDLRGVCVSGCEPIGCDGVLFSTHTLDKCGICQGDGSSC 213
QY 683 KVTGLFTK-PMH-CYNFWFAIPAGASSIDIRQCYKGLIGDNNVLAKNKSGYLLNGH 740
DQ 214 THVTGNRYKNAHLGYSLVTHIPAGARDIOIVERKKA-----DVLADADAGYFFNGN 268
QY 741 FVSAVERDLVVGSLRYST---GTAVESLQASRPILPLTVEVLSV-GKMTPPRVR 795
DQ 269 KYVDS-PKNFNIAGTVVYRRPMDVYETGIEYIVAQGTNOGLNVMVWQNGK--SPSIT 325
QY 796 YSFYLPKPREBDS-----809
DQ 326 FEYTLQPHESRQPIYGFSESAESQGLDGAGLMGFIPHNGSLYQOASERLGLDNL 385
QY 810 -SHP-----812
DQ 386 FGHPLDMELGPSQOETNEVCEQAGGACBPPRGKGRDRNVGTGTLTKDKDEVD 445
QY 813 -----KD-----PR-----816
DQ 446 HFASQEFPSANASDQLLAGSGLDKFTLNETVNSIFAQAPRSSLAESFFVDEENEGA 505
QY 817 GPSVLHNSVLS-----NOVEQPDDRPPA-----841
DQ 506 GPYLLNGSYLELSSDRVANSSEAPPNVSTSLTSAGNRTHKARTPKARKQGVSPADM 565
QY 842 -----R 842
DQ 566 YRWKLSHSPCSATCTTGVMASAYAMCVRYDGEVDDSYCDALTRPEPVHEFCAGREQPR 625
QY 843 WVAGSWGPCASCSGGLQKRAV-----864
DQ 636 WETSSWSECSRTCEGYQFVVRVWKMSPGDFSSVSLCEAAEAVRPEREKTCTRNPA 685
QY 865 -----DCRGSAGRTV-----PACDAARHPVETQAC-GEPCP-TWELSA 901
DQ 686 GPQEMSEWSECTAKGERSVVTDRIRCSDEKLCDDNTRPVGEKNCTGPPCDRQWTVSD 745
QY 902 WSPCKSGGGRGFRSLKCVCHGGRLLARQCNLHRYQBELDFCVLRPC 950
DQ 746 WGPCSGCGGORTIRHVIKTSYSDGRVVPESQCOMETKPLAIHPCGDKNC 794

RESULT 11

T15976
hypothetical protein F08C6.1 - Caenorhabditis elegans
A:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15976
R:Bentley, D.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid F08C6.
A:Reference number: Z18440
A:Accession: T15976
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-957 <BEN>
A:Cross-references: EMBL:U29378; NID:9868184; PID:9868185; PIDN:AAA68721.1; CESP:F08C6.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F08C6.1
A:Introns: 23/1; 135/3; 220/3; 285/3; 325/2; 376/3; 461/2; 498/3; 532/3; 627/2; 714/1; 7

Query Match 8.7%; Score 450.5; DB 2; Length 957;
Best Local Similarity 22.7%; Pred. No. 1.1e-22;
Matches 227; Conservative 106; Mismatches 328; Indels 339; Gaps 47;
QY 57 ITAFQEDFYLHLPDAQFLAPAFSTHGLVPLGGLTGGSSD-----LRRFYSGDVNAEPD 112
DQ 61 IQAFNKYINLSL---EXTLAKLLSSGVTVVKKNEKGGSLDGFSTLDSCHYH---HYGEK 114
QY 113 SFAVSLCCGGL-----RGAFGVYGAEEYVISPLPNASAPAAOANSQ-GAHLLOREG 161
DQ 115 VYAAISNCGRIVSESNRKGKGTVIDDGEIIVVHFPPDHHAHRSRATENGAAHVYKRE 174
QY 162 VPGGSGDPTSRGCVAGWNPAILRALDPYKPRRAGFESRRSRGRKRFRVSIPIR--- 218
DQ 175 T---LAGEPKDFCGLDN-----VVTESLVEDESAIFEDV-----FVTGQRLTQ 215
QY 219 ---YVETLVVADESVMKF---HG--AD--LEHYLLTLTAARLYRHPSTILNINIVY 266
DQ 216 QSLIVELAVFVDENLWRHFSKHHGMADKKQDYTLTLNINQIMYQTPASPPLTFVR 275
QY 267 VKVLLLRDRSG---PKVTGNAALTNRFCAMQKKLNKVSOKHPEYWDATILFTQDL- 321
DQ 276 IRYEVLTRPSALAGYLHNHNAQMYLDRFCRYORN-AYRD-----WDHAIMLTGYDIH 329
QY 322 CGATTCDTLGMADVGTMDCKPSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKKVCEVF 381
DQ 330 RGAGRSISGIAKLDGMDPWNCTLAEGLDFTSAGFIGTHELGHRTFLKDH----- 381
QY 382 GKLRANHMSPTLIQIDRANPWSACSAIITDFLDSGHGDCLLDQPSKPISLPEDLPAS 441
DQ 382 -----KSDTLF-----387
QY 442 YTLSSQCELAFGVGSKPCPYMQYCKLWCTGRAKGMVQCQTRHFWADGTSCGEKLC 501
DQ 388 -----TGTFCGSPKWCOL 400
QY 502 GACVERHNLNK-----HRVDGSAKWDYPGCSR-TCGG-----GVQLA 539
DQ 401 GRCPWPTGNETQPTVOHVAVPVVITLPSRIDGWSGWA-TICSOCTCNGLILSGVLAIA 459
QY 540 RROCTNPTPANGKYCEGVRYKYSCHNLEPCPSASGSKFREECQ-----EAFNGY 590
DQ 460 RRTCSAPYPANGSDCVGSTRAVLCSROCGRASKSVDEYISDKCMEQKRLKNDRELTKG 519
QY 591 NHSTNRLTLAVAWPKYSGVSPRDKCKLICRAN---GTGYFY-VLAPKVVDTGLCSPDST 646
DQ 520 GSQNLNF-----FORACKVPCDVQOQHYGSQNRNRYRFFGDLNPDGTCGYD-- 563
QY 647 SYCVQCKCIKAGCDGNL-----GSKKRFDC-KVCVGD 678
DQ 564 RYCLDGECLALNCNNALLSRDSCPTDCTPITDOSSSVYRQGWGTSLWTSCATCGG 623
QY 679 ----NKCKKVTLTKPMHGYNFVVAIPAGASSIDIRQGYKGLIGDNNVLAKNKSGK 734
DQ 624 YRKNRAC-SITG-----QCEGNEDETEVCSESCPSV-----LRVGNWST 664
QY 735 YLLNGHFVYSAVERDLVWKGSLRYSGTGTAVESLQASRPILPEP---LTVEVLSVGKWT 791
DQ 665 WTEWNHCVSC-----GRGSQARY-----RKCLSPHRTLAFDCPGENKVT- 704
QY 792 PVRYSFYLPKPREDKSHKPRGSPSVLHNSVLSNNOVEQPDDRPP-----ARVWA 845
DQ 705 NELRITFF-----KARSYIMCSVRCNKIKRNTISEKNIEVRSCDNGPCNAIGVWGTW- 756
QY 846 GSWGPCSACSGS---LQKRAVD---CRGSAGORTVPACDAARHPVETQACPEPT---W 897
DQ 757 GWSWTCSTSGGTILVRORTCNREPCDSAHERR--SCNVA-----TCQNDGIW 803
QY 898 EL-SAWSPCSCGGRGFORRSKLVCHG--GRLLARDQCN 934
DQ 804 SLWNWSDCSRVCVGLRSLRSRSCFGSGCMGASSEQFCN 843

Db 379 CGTPENQNECCDA-----ATCKLK-----SGSQCHGD-----CCEQCKFSKSTECR 422
QY 517 GSAKWDYPCSRCTCGGVOLARRCTNPTPANGKYGCEGVRYKRSNLEPCP-----571
Db 423 ASMSCDPAHCFTGQS-----ECPADVFNKNGQPCLD---NYGYCNGNCPTIMYHQ 471
QY 572 -----SSASGKSFREEOCAFNGYNHSTNRLTLAVAMPKYSGVSPRD-KC-KLIC 620
Db 472 CYALFGADVYEADSCFKDNQNGYGYCKRNGKIKPCA-----PEDVKGRLYC 522
QY 621 RANGTG-----YFYVLAPK-----VVDGTLCSPDSTSVCVQKCI 655
Db 523 KDNSPGQNNPKMFYNDDEHKGMVLPGTKA--DGKVCNSGHCY 565
RESULT 14
S53270
catrocollastatin precursor - western diamondback rattlesnake
C:Species: Crotalus atrox (western diamondback rattlesnake)
C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C:Accession: S55270
R:Zhou, Q.; Smith, J.B.; Grossman, M.H.
Biochem. J. 307, 411-417, 1995
A:Title: Molecular cloning and expression of catrocollastatin, a snake-venom protein from
A:Reference number: S55264; MUID:95251603; PMID:7733877
A:Accession: S55270
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-609 <ZHO>
A:Cross-references: GB:U21003; NID:9710353; PIDN:AAC59672.1; PID:g710354
C:Superfamily: mouse meltrin alpha; disintegrin homology

Query Match 6.5%; Score 337; DB 2; Length 609;
Best Local Similarity 22.8%; Pred. No. 3.5e-15;
Matches 150; Conservative 90; Mismatches 259; Indels 160; Gaps 30;

QY 50 DQGLIFQITAFQEDFYLLHTPDQAFLAPAFSTEHLGVPLQGLTGSSDLRRFCYSGDVNA 109
Db 52 EDAMQYELKVNQEPVVLHLGKNGKGLFSKDYSETHYSPDGREITTYPLVEDHCYHGRITEN 111
QY 110 EPDSFAAVSLGGLRGAFGCGAEYVISP--LPNASAPAAQNSOGAHLQRRGVPGGFS 167
Db 112 DADSTASISACNGLKGFHKLQEGMLIEPLKPLSEAHAVKYENVE-----KE 160
QY 168 GPTSRGCVASGNPAILRALDPYKPRAGFGESRRSRRAKRFVSP-----RY 219
Db 161 DEALKMCGVTQNW-----ESYEPIK-----KASQIVTAHQKYNPRF 199
QY 220 VETLVVADESVMKFGADLEHY---LLTLATAALYRHPSTILNPINIVVVKLLRRDRD 276
Db 200 VELFLVVDKAMVTKNNGDLDKIKTRMYEIVNTVNEYRYMY---IHVALVGLSEIWSNED 255
QY 277 SGPKVT--GNAALTILRNFCAMOKKLNKYSKHPYWDTAILFTRODLGATTCDTLGMAD 334
Db 256 ---KIIVKEPAGVTNAGFEWARKTDLTRKKH-----DNAQLTAIDL-----DRVIGLAY 303
QY 335 VETMCDPKRSCSVIED---DGLPSAFTTAHELGHVFNPHDN-VKCEEVFGKLRANHMM 390
Db 304 VGSMPKPKRSTGIQDYSEINLVAVIMAHENGHNINHDSGYSCSGDYACIMRPEISP 363
QY 391 SPTLIQIDRANPWSACSAAITDFLDSHGDCLLDQ-----SKPISLPEDLPASVTL 444
Db 364 EPSTF-----FSNCSYFECWDFIMNHNPECILNEPLGTDIISPPVCGNELL-----EV 411
QY 445 SQCELAFLGCVGSKPCYMOYCTKLTCTGKAKQMWYQCTHFPWADGTCGEGKCLKLGAC 504
Db 412 GECED-----CGTPENCQNECCDA-----ATCKLK-----SGSQCHGD-----C 446
QY 505 VERHNLNKRHVD--GSWAKWDYPCSRCTCGGVOLARRCTNPTPANGKYGCEGVRYKY 562
Db 447 CEQCKFSKSTECRASMSCEDPAEHC TGQS-----ECPADVFNKNGQPCLD---NY 495

QY 563 RSCNLEPCP-----SSASGKSFREEOCAFNGYNHSTNRLTLAVAWVPKYS 608
Db 496 GYCYNGNCPIIMYHOCYDILFGADVYEADSCFERNOKNGYGYCKRNGKNIKPCA-----549
QY 609 GVSPRD-KC-KLICRANGTG-----YFYVLAPK-----VVDGTLCSPDSTSVCVQKCI 655
Db 550 ---PEDVKGRLYCKDNSPGQNNPKMFYNDDEHKGMVLPGTKA--DGKVCNSGHCY 603
RESULT 15
S60257
meltrin alpha - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 20-Jun-2000
C:Accession: S60257
R:Yagami-Hiromasa, T.; Sato, T.; Kurisaki, T.; Kamiyo, K.; Nabeshima, Y.I.; Fujisawa
Nature 377, 652-656, 1995
A:Title: A metalloprotease-disintegrin participating in myoblast fusion.
A:Reference number: S60257; MUID:96026308; PMID:7566181
A:Accession: S60257
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-903 <YAG>
A:Cross-references: EMBL:D50411; NID:g1054586; PIDN:BAA08912.1; PID:g1054587
C:Superfamily: mouse meltrin alpha; disintegrin homology
F:421-503/Domain: disintegrin homology <DIS>
F:349/Active site: Glu #status predicted

Query Match 6.4%; Score 329; DB 2; Length 903;
Best Local Similarity 21.5%; Pred. No. 2e-14;
Matches 159; Conservative 74; Mismatches 245; Indels 262; Gaps 27;

QY 2 LLGLTLLAPAGRTA-----GGPEPEREVVPIRLDPDINGRYYRWGPEDSGDQ 52
Db 15 LLLALAGALLAPRAAGMSLWDQGA YEVARSL--LSKDPGIPQGI-----PAKHPDV 68
QY 53 LIFOITAFQEDFYLLHTPDQAFLAPAFSTEHLGVPLQ-----GLTGGSSDLRRFCYSGDV 107
Db 69 LTVQLQLESRDLLSLERNEGLIANGFTETHY---LQDGTDVSLTRNHTD--HCYYHGHV 123
QY 108 NAEPSFAAVSLGGLRGAFGCGAEYVISP--LPNAS-----APA-AORNQSGAHLQRRG 161
Db 124 QGDAASVVSSTCSDLRGLIMFENKTYSLPEMKNTTDSYKLVPAESMTNIQGL-----176
QY 162 VPGGPGDPTSRGCVASGNPAILRALDPYKPRAGFGESRRSRRAKRFVSPYVE 221
Db 177 -----CGSQHNKSNLTWEDVSP-----GTSQMRARHK---RETUKMTIYE 215
QY 222 TLVVADESVMKFGADLEHYLLTLATAALYRHPSTILNPINIVVVKLLRRDRDSGPV 281
Db 216 LVIADVADREFORQKDLKVKQRLIETANHVDFKFPRLN-IRIVLVGVEVWNDIDK--CSI 273
QY 282 TGNAAITILRNFCAMOKKLNKYSKHPYWDTAILFTRODLGCGATTCDTLGMADVGMCDP 341
Db 274 SQDPFTRLHEFLDWRKTKLLPRKSH---DNAQLISGYFQGT---TIGMAPIMSCTA 325
QY 342 KRSCSVI---EDDGLPSAFTTAHELGHVFNPHDN-VKCEEVFGKLRANHMSPTLIQI 397
Db 326 EQSGVVDHSDSPGAAVTLAHELGHVFNPHDNHDTLBERGCSRMMAEKGCGIMNPS-----381
QY 398 DRANP-----WSACSAAITDFLDSHGDCLLDQ-----SKPISLPEDLPAS---427
Db 382 -TGFPFVWFVSSCRKDLASLEKGMGMCFLNLPVQKAFGGRKCGNGYVEEGECDCGE 440
QY 428 -----YTLSSQCELAFLGCVGSKPCYMOY-----SKPISLPEDLPAS---441
Db 441 PECTNRCNATTCTLKPDVCAHGOCCEDCOLKPPGTACRGSNSCDLPEFCTGAPHC 500
QY 442 -----YTLSSQCELAFLGCVGSKPCYMOY-----SKPISLPEDLPAS---464
Db 501 PANYLHDGHPCCQVDCYCYNGICQTHQEOCVTLWGPQAKPAPGICFERVNSAGDPYGN 560
QY 465 -----CTKLWCTGKAK-----GOMVQTRHFPW 487

```
Db 561 GKDSKSAFAKELRDAKCGKIQCQGGASRPVIGTNAVSIETNIPQOEGGRILCRGTHVYL 620
Qy 488 AD-----GTSCEGKCLKAGACVERHNLNKHVRVDSWAKWDPYGPCSRTC--GGGV 536
Db 621 GDDMPDPGLVLAGTKCAEGKICLNRRQCNTISVFGVHK-----CAMQCHGRGV 667
Qy 537 QLARROCTNPTPANGGKYCE 556
Db 668 CNRRKNC-----HCE 677
```

Search completed: April 29, 2003, 17:13:17
Job time : 44.9736 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:22:38 ; Search time 26 Seconds
(without alignments)
2927.820 Million cell updates/sec

Title: US-10-009-332-1
Perfect score: 950
Sequence: 1 MLLIGILTLAFAGTAGGFE.....DQCNLHRKPQLDFCVLRPC 950

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 301932 seqs, 80129803 residues

Word size : 0

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pap.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pap.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	931	98.0	950	10 US-09-965-631-4	Sequence 4, Appli
2	408	42.9	823	9 US-10-163-316-2	Sequence 2, Appli
3	345	36.3	367	10 US-09-965-631-6	Sequence 6, Appli
4	300	31.6	321	10 US-09-965-631-2	Sequence 2, Appli
5	17	1.8	608	10 US-09-803-589-2	Sequence 2, Appli
6	17	1.8	608	10 US-09-803-589-8	Sequence 8, Appli
7	17	1.8	608	12 US-10-105-929-13	Sequence 13, Appli
8	17	1.8	727	9 US-10-097-597-1	Sequence 1, Appli
9	17	1.8	727	9 US-10-097-597-12	Sequence 12, Appli
10	17	1.8	727	9 US-10-097-580-1	Sequence 1, Appli
11	17	1.8	727	9 US-10-097-580-12	Sequence 12, Appli
12	17	1.8	727	10 US-09-445-023A-1	Sequence 1, Appli
13	17	1.8	727	10 US-09-445-023A-12	Sequence 12, Appli
14	17	1.8	950	10 US-09-321-987B-4	Sequence 4, Appli
15	17	1.8	967	12 US-10-105-929-2	Sequence 2, Appli
16	17	1.8	968	9 US-10-163-316-7	Sequence 7, Appli
17	14	1.5	518	10 US-09-803-589-10	Sequence 10, Appli
18	14	1.5	551	12 US-10-105-929-16	Sequence 16, Appli
19	13	1.4	837	9 US-10-174-590-352	Sequence 352, App

20	13	1.4	837	9	US-10-176-758-352	Sequence 352, App
21	13	1.4	837	9	US-10-175-737-352	Sequence 352, App
22	13	1.4	837	9	US-10-173-706-352	Sequence 352, App
23	13	1.4	837	9	US-10-175-738-352	Sequence 352, App
24	13	1.4	837	9	US-10-175-752-352	Sequence 352, App
25	13	1.4	837	9	US-10-176-482-352	Sequence 352, App
26	13	1.4	837	9	US-10-176-757-352	Sequence 352, App
27	13	1.4	837	9	US-10-176-913-352	Sequence 352, App
28	13	1.4	837	9	US-10-180-552-352	Sequence 352, App
29	13	1.4	837	9	US-10-180-557-352	Sequence 352, App
30	13	1.4	837	9	US-10-173-700-352	Sequence 352, App
31	13	1.4	837	9	US-10-174-572-352	Sequence 352, App
32	13	1.4	837	9	US-10-174-579-352	Sequence 352, App
33	13	1.4	837	9	US-10-174-582-352	Sequence 352, App
34	13	1.4	837	9	US-10-174-588-352	Sequence 352, App
35	13	1.4	837	9	US-10-175-739-352	Sequence 352, App
36	13	1.4	837	9	US-10-175-740-352	Sequence 352, App
37	13	1.4	837	9	US-10-175-743-352	Sequence 352, App
38	13	1.4	837	9	US-10-176-488-352	Sequence 352, App
39	13	1.4	837	9	US-10-176-492-352	Sequence 352, App
40	13	1.4	837	9	US-10-176-747-352	Sequence 352, App
41	13	1.4	837	9	US-10-176-750-352	Sequence 352, App
42	13	1.4	837	9	US-10-176-985-352	Sequence 352, App
43	13	1.4	837	9	US-10-176-987-352	Sequence 352, App
44	13	1.4	837	9	US-10-176-991-352	Sequence 352, App
45	13	1.4	837	9	US-10-176-992-352	Sequence 352, App

ALIGNMENTS

RESULT 1
US-09-965-631-4
; Sequence 4, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: NO. US20020115842A1el Human Proteases and Polynucleotides Enco
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-965-631-4

Query Match 98.0%; Score 931; DB 10; Length 950;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 931; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	20	EPERVVVPIRLDPDINGRRYYWRGPGSDGGLFQITAFQEDFYHLHTPDQAFAPAF 79	
DB	20	EPERVVVPIRLDPDINGRRYYWRGPGSDGGLFQITAFQEDFYHLHTPDQAFAPAF 79	
QY	80	STHLGVPLQGLTGSSDLRCFYSGDVNABPDSFAAVSLCGGLRGAFYRGAEVVISPL 139	
DB	80	STHLGVPLQGLTGSSDLRCFYSGDVNABPDSFAAVSLCGGLRGAFYRGAEVVISPL 139	
QY	140	PNASAPAAQRNSQGAHLQRGVGPGSDPTSRGCVASGWNPAIRALDYPKPRRAGFG 199	
DB	140	PNASAPAAQRNSQGAHLQRGVGPGSDPTSRGCVASGWNPAIRALDYPKPRRAGFG 199	
QY	200	ESRRRRSGRAKRFVSIPIRYVETLVVADESMVKFHGADLEHYLLTATARLYRHSIL 259	
DB	200	ESRRRRSGRAKRFVSIPIRYVETLVVADESMVKFHGADLEHYLLTATARLYRHSIL 259	
QY	260	NPINIVVYKVLRLDRDSGPVKTGNAALTLRNFCAWOKLKNKVDKHPYWDTAILFTQ 319	

```
|||||
Db 260 NPINIVVVKVLLLRDRSGPKVTGNAALTLRNFCAWQKLNKVSQKHPYWDTAILFTQ 319
QY 320 DLGATTCDTLGMADVGTMC DPRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEE 379
Db 320 DLGATTCDTLGMADVGTMC DPRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEE 379
QY 380 VFGKLRANHHMSP TLIQIDRANPWSACSAIITDFLDSHGDCLLDQPSKPIISLPEDLP 439
Db 380 VFGKLRANHHMSP TLIQIDRANPWSACSAIITDFLDSHGDCLLDQPSKPIISLPEDLP 439
QY 440 ASYTLSSQCELAFGVSGKPCPYMQYCTKLWCTGKAGQMVQCOTRHPFWADGTSCGEGKLC 499
Db 440 ASYTLSSQCELAFGVSGKPCPYMQYCTKLWCTGKAGQMVQCOTRHPFWADGTSCGEGKLC 499
QY 500 LKACVERHNLNHRVDGSAWAKWDPYPCSRCTCGGGVQLARRQCTNPTPANGKCYCEGVR 559
Db 500 LKACVERHNLNHRVDGSAWAKWDPYPCSRCTCGGGVQLARRQCTNPTPANGKCYCEGVR 559
QY 560 VKYBSCNLEPCPSSASGKSPREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLI 619
Db 560 VKYBSCNLEPCPSSASGKSPREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLI 619
QY 620 CRANGTYFYVLA PKVVDGTLCS PDSTSV CVQKCIKAGCDGNLGSKKRDKCGVCGGDN 679
Db 620 CRANGTYFYVLA PKVVDGTLCS PDSTSV CVQKCIKAGCDGNLGSKKRDKCGVCGGDN 679
QY 680 KSCKVTGLFTKPMHGNFVVAIPAGASSIDIRQGYKGLIGDDNYLALNSOGKYLLNG 739
Db 680 KSCKVTGLFTKPMHGNFVVAIPAGASSIDIRQGYKGLIGDDNYLALNSOGKYLLNG 739
QY 740 HFVYSAYERDLVVKGLSLRYSGTAVESLQASRPILPTVEVLVSIGKMTPPRVYSFY 799
Db 740 HFVYSAYERDLVVKGLSLRYSGTAVESLQASRPILPTVEVLVSIGKMTPPRVYSFY 799
QY 800 LPKEPREDKSHPKDPRGSPVLHNSVLSLSNQVEQDDPRPARWAGSWGPCASCSGSL 859
Db 800 LPKEPREDKSHPKDPRGSPVLHNSVLSLSNQVEQDDPRPARWAGSWGPCASCSGSL 859
QY 860 QKRAVDCGSAGORTVPACDAARHPVETOACGECPTWELSAWSPCKSCGRGFQRRSLK 919
Db 860 QKRAVDCGSAGORTVPACDAARHPVETOACGECPTWELSAWSPCKSCGRGFQRRSLK 919
QY 920 CVHGGRLLARDQCNLHRKPKQLDFCVLRPC 950
Db 920 CVHGGRLLARDQCNLHRKPKQLDFCVLRPC 950
```

RESULT 2

```
US-10-163-316-2
; Sequence 2, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; FILE REFERENCE: MP101-025P1RNM
; CURRENT APPLICATION NUMBER: US/10/163.316
; PRIOR FILING DATE: 2002-06-05
; PRIOR FILING DATE: 2002-06-05
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-316-2
```

```
Query Match 42.9%; Score 408; DB 9; Length 823;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 107 VNAEPDSFAAVSLCGGLRGATGYRGAEEVISP LNPASAPAAQRNSQGAHLLQRRGVPGGP 166
Db 107 VNAEPDSFAAVSLCGGLRGATGYRGAEEVISP LNPASAPAAQRNSQGAHLLQRRGVPGGP 166
QY 167 SGDPYSGVAGWNPATLRALDPYKPRRAGFGESESRSSRRSRGAKRFVSPRYVETLVVA 226
Db 167 SGDPYSGVAGWNPATLRALDPYKPRRAGFGESESRSSRRSRGAKRFVSPRYVETLVVA 226
QY 227 DESWYKPHGADLEHYLLTLLATAARLYRHPSILNPINIVVVKVLLLRDRSGPKVTGNA 286
Db 227 DESWYKPHGADLEHYLLTLLATAARLYRHPSILNPINIVVVKVLLLRDRSGPKVTGNA 286
QY 287 LTLNFCNAWQKLNKVSQKHPYWDTAILFTRODLGCGATTCDTLGMADVGTMC DPRKSCS 346
Db 287 LTLNFCNAWQKLNKVSQKHPYWDTAILFTRODLGCGATTCDTLGMADVGTMC DPRKSCS 346
QY 347 VTEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHHMSP TLIQIDRANPWSAC 406
Db 347 VTEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHHMSP TLIQIDRANPWSAC 406
QY 407 SAAIITDFLDSHGDCLLDQPSKPIISLPEDLPASVYTLSSQCELAFGVSGKPCPYMQYCT 466
Db 407 SAAIITDFLDSHGDCLLDQPSKPIISLPEDLPASVYTLSSQCELAFGVSGKPCPYMQYCT 466
QY 467 KLWCTGKAGQMVQCOTRHPFWADGTSCGEGKLC LKACVERHNLNHR 514
Db 467 KLWCTGKAGQMVQCOTRHPFWADGTSCGEGKLC LKACVERHNLNHR 514
```

RESULT 3

```
US-09-965-631-6
; Sequence 6, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Fridge, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucleotides Enco
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; PRIOR FILING DATE: 2001-09-27
; PRIOR FILING DATE: 2001-09-27
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 367
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-965-631-6
```

```
Query Match 36.3%; Score 345; DB 10; Length 367;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 20 EPREVVPVRLDPDINGRRYWRGPEDSGDGLIFQITAFQEDFYHLHTPDQAFLAPAF 79
Db 20 EPREVVPVRLDPDINGRRYWRGPEDSGDGLIFQITAFQEDFYHLHTPDQAFLAPAF 79
QY 80 STEHLGVPLQGLTGGSSDLRCRCFSGDVNAEPDFAVSLCGGLRGATGYRGAEEVISP 139
Db 80 STEHLGVPLQGLTGGSSDLRCRCFSGDVNAEPDFAVSLCGGLRGATGYRGAEEVISP 139
QY 140 PNASAPAAQRNSQGAHLLQRRGVPGGPDSGPTSRCSVAGSWNPAILRALDPYKPRRAGFG 199
Db 140 PNASAPAAQRNSQGAHLLQRRGVPGGPDSGPTSRCSVAGSWNPAILRALDPYKPRRAGFG 199
QY 200 ESKRRSRGAKRFVSPRYVETLVVADESMWKFHGDLEHYLLTLLATAARLYRHPSIL 259
Db 200 ESKRRSRGAKRFVSPRYVETLVVADESMWKFHGDLEHYLLTLLATAARLYRHPSIL 259
QY 260 NPINIVVVKVLLLRDRSGPKVTGNAALTLRNFCAWQKLNKVSQKHPYWDTAILFTQ 319
Db 260 NPINIVVVKVLLLRDRSGPKVTGNAALTLRNFCAWQKLNKVSQKHPYWDTAILFTQ 319
```

QY 320 DLGGATTCITLGMADVGTMCMDPKRSCSVIEDDGLPSAFTTAHELG 364
 Db 320 DLGGATTCITLGMADVGTMCMDPKRSCSVIEDDGLPSAFTTAHELG 364

RESULT 4

US-09-965-631-2
 ; Sequence 2, Application US/09965631
 ; Patent No. US20020115842A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fridde, Carl Johan
 ; APPLICANT: Hilbun, Erin
 ; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Encodin
 ; FILE REFERENCE: LEX-0241-USA
 ; CURRENT APPLICATION NUMBER: US/09/965,631
 ; CURRENT FILING DATE: 2001-09-27
 ; PRIOR APPLICATION NUMBER: US 60/236,689
 ; PRIOR FILING DATE: 2000-09-29
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 321
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 US-09-965-631-2

Query Match 31.6%; Score 300; DB 10; Length 321;
 Best Local Similarity 100.0%; Pred. No. 4.4e-279;
 Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 EPERVVPRIILDPDINGRRYWRGPDSDGQGLIFQITAFQEDFYHLTTPDAQFLAPAF 79
 Db 20 EPERVVPRIILDPDINGRRYWRGPDSDGQGLIFQITAFQEDFYHLTTPDAQFLAPAF 79
 QY 80 STEHLGVPLQGTGSSDLRCFYSGDVNAEPDSPAAYSLCGLRGAFYGAEYVISPL 139
 Db 80 STEHLGVPLQGTGSSDLRCFYSGDVNAEPDSPAAYSLCGLRGAFYGAEYVISPL 139
 QY 140 PNASAPAAQNSQGAHLQVRGVPDGPSPDPTSRGCVASGNPAILRALDYPKPRAGFG 199
 Db 140 PNASAPAAQNSQGAHLQVRGVPDGPSPDPTSRGCVASGNPAILRALDYPKPRAGFG 199
 QY 200 ESRRRRSRAKRFYSIPRYVETLVVADESVMVKFHGADLEHLLTLATAARLYRHPISIL 259
 Db 200 ESRRRRSRAKRFYSIPRYVETLVVADESVMVKFHGADLEHLLTLATAARLYRHPISIL 259
 QY 260 NPINIVVVKVLLLRDSDGPKVTGNAALTNRNCAWQKKLNKSDKHPEYWDTAILFTTRQ 319
 Db 260 NPINIVVVKVLLLRDSDGPKVTGNAALTNRNCAWQKKLNKSDKHPEYWDTAILFTTRQ 319

RESULT 5

US-09-803-589-2
 ; Sequence 2, Application US/09803589
 ; Patent No. US20020112251A1
 ; GENERAL INFORMATION:
 ; APPLICANT: McCarthy, Sean A.
 ; APPLICANT: Holtzman, Douglas A.
 ; APPLICANT: Goodearl, Andrew D.J.
 ; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
 ; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
 ; FILE REFERENCE: 07334-325001
 ; CURRENT APPLICATION NUMBER: US/09/803,589
 ; CURRENT FILING DATE: 2001-03-09
 ; PRIOR APPLICATION NUMBER: US 09/128,709
 ; PRIOR FILING DATE: 1998-08-04
 ; PRIOR APPLICATION NUMBER: US 60/054,645
 ; PRIOR FILING DATE: 1997-08-04
 ; PRIOR APPLICATION NUMBER: US 09/130,491
 ; PRIOR FILING DATE: 1998-08-06
 ; PRIOR APPLICATION NUMBER: US 60/054,966

; PRIOR FILING DATE: 1997-08-06
 ; PRIOR APPLICATION NUMBER: US 60/058,108
 ; PRIOR FILING DATE: 1997-09-05
 ; PRIOR APPLICATION NUMBER: US 09/388,280
 ; PRIOR FILING DATE: 1999-09-01
 ; PRIOR APPLICATION NUMBER: US 09/388,279
 ; PRIOR FILING DATE: 1999-09-01
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 608
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-803-589-2

Query Match 1.8%; Score 17; DB 10; Length 608;
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
 Db 91 AFTTAHELGHVFNMPHD 107

RESULT 6

US-09-803-589-8
 ; Sequence 8, Application US/09803589
 ; Patent No. US20020112251A1
 ; GENERAL INFORMATION:
 ; APPLICANT: McCarthy, Sean A.
 ; APPLICANT: Holtzman, Douglas A.
 ; APPLICANT: Goodearl, Andrew D.J.
 ; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
 ; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
 ; FILE REFERENCE: 07334-325001
 ; CURRENT APPLICATION NUMBER: US/09/803,589
 ; CURRENT FILING DATE: 2001-03-09
 ; PRIOR APPLICATION NUMBER: US 09/128,709
 ; PRIOR FILING DATE: 1998-08-04
 ; PRIOR APPLICATION NUMBER: US 60/054,645
 ; PRIOR FILING DATE: 1997-08-04
 ; PRIOR APPLICATION NUMBER: US 09/130,491
 ; PRIOR FILING DATE: 1998-08-06
 ; PRIOR APPLICATION NUMBER: US 60/054,966
 ; PRIOR FILING DATE: 1997-08-06
 ; PRIOR APPLICATION NUMBER: US 60/058,108
 ; PRIOR FILING DATE: 1997-09-05
 ; PRIOR APPLICATION NUMBER: US 09/388,280
 ; PRIOR FILING DATE: 1999-09-01
 ; PRIOR APPLICATION NUMBER: US 09/388,279
 ; PRIOR FILING DATE: 1999-09-01
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 608
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-09-803-589-8

Query Match 1.8%; Score 17; DB 10; Length 608;
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
 Db 91 AFTTAHELGHVFNMPHD 107

RESULT 7

US-10-105-929-13
 ; Sequence 13, Application US/10105929
 ; Patent No. US20020137142A1

```
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: US/10/105,929
; PRIOR FILING DATE: EARLIER APPLICATION NUMBER: 09/130,491
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 13
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-105-929-13

Query Match 1.8%; Score 17; DB 12; Length 608;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
Db 91 AFTTAHELGHVFNMPHD 107
|||||

RESULT 8
US-10-097-597-1
; Sequence 1, Application US/10097597
; Publication No. US20030022352A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiiji
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: Pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-597-1

Query Match 1.8%; Score 17; DB 9; Length 727;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
Db 156 AFTTAHELGHVFNMPHD 172
|||||

RESULT 9
US-10-097-597-12
; Sequence 12, Application US/10097597
; Publication No. US20030022352A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiiji
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: Pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-097-597-12

Query Match 1.8%; Score 17; DB 9; Length 727;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
Db 156 AFTTAHELGHVFNMPHD 172
|||||

RESULT 10
US-10-097-580-1
; Sequence 1, Application US/10097580
; Publication No. US20030032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiiji
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-580-1

Query Match 1.8%; Score 17; DB 9; Length 727;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
Db 156 AFTTAHELGHVFNMPHD 172
|||||
```


RESULT 11
US-10-097-580-12
; Sequence 12, Application US/10097580
; Publication No. US2003032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-097-580-12

Query Match 1.8%; Score 17; DB 9; Length 727;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
|||||
DB 156 AFTTAHELGHVFNMPHD 172

RESULT 12
US-09-445-023A-1
; Sequence 1, Application US/09445023A
; Patent No. US20020119167A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/09/445,023A
; CURRENT FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-445-023A-1

Query Match 1.8%; Score 17; DB 10; Length 727;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
|||||
DB 156 AFTTAHELGHVFNMPHD 172

RESULT 13
US-09-445-023A-12
; Sequence 12, Application US/09445023A
; Patent No. US20020119167A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human AD
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/09/445,023A
; CURRENT FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-445-023A-12

Query Match 1.8%; Score 17; DB 10; Length 727;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
|||||
DB 156 AFTTAHELGHVFNMPHD 172

RESULT 14
US-09-321-987B-4
; Sequence 4, Application US/09321987B
; Patent No. US20020102210A1
; GENERAL INFORMATION:
; APPLICANT: Bielloch, Robert H
; TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
; FILE REFERENCE: 960296.95386
; CURRENT APPLICATION NUMBER: US/09/321,987B
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087,170
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/129,023
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Murine
US-09-321-987B-4

Query Match 1.8%; Score 17; DB 10; Length 950;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
|||||
DB 380 AFTTAHELGHVFNMPHD 396

RESULT 15
US-10-105-929-2
; Sequence 2, Application US/10105929
; Patent No. US20020137142A1

```

; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/10/105,929
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/130,491
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-105-929-2

Query Match      1.8%; Score 17; DB 12; Length 967;
Best Local Similarity 100.0%; Pred. No. 1.7e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 356 AFTTAHELGHVFNMPHD 372
    |
Db 396 AFTTAHELGHVFNMPHD 412

Search completed: April 29, 2003, 17:25:35
Job time : 31 secs

```

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 45.0027 seconds
(without alignments)
3378.970 Million cell updates/sec

Title: US-10-009-332-1_copy_213_950
Perfect score: 4043
Sequence: 1 FVSIPRYVETLVVADESMWK.....DQCNLHRKQELDFCVLRPC 738

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mhc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.rodent.*
- 12: sp.virus.*
- 13: sp.vertebrate.*
- 14: sp.unclassified.*
- 15: sp.virus.*
- 16: sp.bacteriap.*
- 17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4043	100.0	950	4 Q8TE58	Q8TE58 homo sapien
2	1685	41.7	340	11 Q91256	Q91256 mus musculus
3	1567	38.8	2165	5 Q19791	Q19791 caenorhabdi
4	1152	28.5	1688	5 Q8SX80	Q8SX80 drosophila
5	1143	28.3	1081	4 Q8TE60	Q8TE60 homo sapien
6	1139.5	28.2	1072	4 Q8TE57	Q8TE57 homo sapien
7	1119.5	27.7	1229	5 Q9VF61	Q9VF61 drosophila
8	1117	27.6	1054	5 Q9W493	Q9W493 drosophila
9	1057	26.1	1095	4 Q8TE56	Q8TE56 homo sapien
10	1024	25.3	1223	4 Q8WX58	Q8WX58 homo sapien
11	1022.5	25.3	1159	4 Q8TEY8	Q8TEY8 homo sapien
12	1020	25.2	1223	4 Q8TE55	Q8TE55 homo sapien
13	1012	25.0	1207	4 Q8TE59	Q8TE59 homo sapien
14	865	21.4	1427	4 Q96L37	Q96L37 homo sapien
15	830	20.5	269	6 Q9GL54	Q9GL54 oryctolagus
16	706	17.5	1280	11 Q9EPX2	Q9EPX2 mus musculus

17	693	17.1	1235	4	O95428	homo sapien
18	681	16.8	2167	5	O76840	caenorhabdi
19	663.5	16.4	1572	5	O44938	haemochus
20	581.5	14.4	2174	5	O9G0R0	drosophila
21	579.5	14.3	3060	5	O9VAV4	drosophila
22	569.5	14.1	766	4	P82987	homo sapien
23	558	13.8	192	6	O95N24	equus caball
24	557	13.8	3198	5	O9UG88	manduca sex
25	553	13.7	1062	5	O19204	caenorhabdi
26	550.5	13.6	525	4	O96RW4	homo sapien
27	518	12.8	187	6	O95N23	equus caball
28	510.5	12.6	872	5	O22580	caenorhabdi
29	504	12.5	1444	5	Q17591	caenorhabdi
30	493	12.2	951	4	O60345	homo sapien
31	487.5	12.1	1014	5	O95R33	drosophila
32	429	10.6	1091	5	O9W126	drosophila
33	426	10.5	790	5	O8T458	drosophila
34	411.5	10.2	1059	5	P90884	caenorhabdi
35	410	10.1	183	11	O99JP1	mus musculus
36	402.5	10.0	364	4	O9UGQ1	homo sapien
37	343	8.5	622	4	O9H8X0	homo sapien
38	338	8.4	192	11	O9CX59	mus musculus
39	332.5	8.2	105	6	O9GLK7	oryctolagus
40	321.5	8.0	244	11	O9CRC7	mus musculus
41	303	7.5	108	6	O9GLK6	oryctolagus
42	302.5	7.5	1637	6	O9XSV8	bos taurus
43	302.5	7.5	5146	6	O8SPM4	bos taurus
44	298	7.4	117	11	O8VHK4	mus musculus
45	279.5	6.9	1081	5	O9U631	drosophila

ALIGNMENTS

RESULT 1

Q8TE58	PRELIMINARY;	PRT;	950 AA.
ID	Q8TE58		
AC	Q8TE58;		
DT	01-JUN-2002 (TRENBLrel. 21, Created)		
DT	01-JUN-2002 (TRENBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TRENBLrel. 21, Last annotation update)		
DE	Metalloprotease disintegrin 15 with thrombospondin domains.		
GN	ADAMTS15.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21856482; PubMed=11867212;		
RA	Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,		
RA	Lopez-otin C.		
RT	"Cloning, expression analysis, and structural characterization of		
RT	seven novel human ADAMTS, a family of metalloproteinases with		
RT	disintegrin and thrombospondin-1 domains.*;		
RL	Gene 283:49-62(2002).		
DR	EMBL; AJ315733; CAC86014.1; .		
KW	Integrin; Protease.		
SQ	SEQUENCE 950 AA; 103286 MW; 5DFBE18285CCCC3B CRC64;		
Query Match	100.0%; Score 4043; DB 4; Length 950;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 738;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 FVSIPRYVETLVVADESMWKFGADLEHYLLTLATAARLYRHPSILNPINVVVKVLL 60		
Db	213 FVSIPRYVETLVVADESMWKFGADLEHYLLTLATAARLYRHPSILNPINVVVKVLL 272		
QY	61 RDRDSGPKVTGNAALTNRNFCWQKLNKVSQKPEYWDTAILFTRODLGATTCDFLGM 120		
Db	273 RDRDSGPKVTGNAALTNRNFCWQKLNKVSQKPEYWDTAILFTRODLGATTCDFLGM 332		
QY	121 ADVGTMCDPKSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKCEVFGKLRANHMSP 180		

```
Db 333 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVYKCEEVFGKLRANHMSP 392
QY 181 TLTIQIDRANPWSACSAAIITDPLDSHGSDCLLDOPSKPISLPEDLPASYYTLISQOCELAF 240
Db 393 TLTIQIDRANPWSACSAAIITDPLDSHGSDCLLDOPSKPISLPEDLPASYYTLISQOCELAF 452
QY 241 GVGSRPCPTMYQCTKLWCTGKRAKGMVQTRHFWADGTSCEGKGLCLKACVRRHNLNK 300
Db 453 GVGSRPCPTMYQCTKLWCTGKRAKGMVQTRHFWADGTSCEGKGLCLKACVRRHNLNK 512
QY 301 HRVDGSAKWDYPGCSRTCCGGVGLARRQCTNPTPANGKYCGVRVYKRSCLNLEPCPS 360
Db 513 HRVDGSAKWDYPGCSRTCCGGVGLARRQCTNPTPANGKYCGVRVYKRSCLNLEPCPS 572
QY 361 SASGKSFFREECEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKKLCIKRANGTGYFYVLA 420
Db 573 SASGKSFFREECEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKKLCIKRANGTGYFYVLA 632
QY 421 PKVVDGTLCSPDSTSVCVQKCIKAGCDNGLSKRRDKPCGVCGDNKSCKKVTGLFTKP 480
Db 633 PKVVDGTLCSPDSTSVCVQKCIKAGCDNGLSKRRDKPCGVCGDNKSCKKVTGLFTKP 692
QY 481 MHGYNFVVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQGYLLNGHFVYVSAVERDLVY 540
Db 693 MHGYNFVVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQGYLLNGHFVYVSAVERDLVY 752
QY 541 KGSLLRYSGTGTAVESLOASRPILPLAVEVLSVGMKTPPRVYSFYLPKPREPREDKSSHP 600
Db 753 KGSLLRYSGTGTAVESLOASRPILPLAVEVLSVGMKTPPRVYSFYLPKPREPREDKSSHP 812
QY 601 KDPGSPVLHNSVLSNQVEQDPPRPPARVWAGSWGPCSCASGSGQKRAVDCRGSAQ 660
Db 813 KDPGSPVLHNSVLSNQVEQDPPRPPARVWAGSWGPCSCASGSGQKRAVDCRGSAQ 872
QY 661 RTVPACDAARHVPETQACGECPTWELSAWSPCSCSGRGFQRLSKVCGHGGRLARDQ 720
Db 873 RTVPACDAARHVPETQACGECPTWELSAWSPCSCSGRGFQRLSKVCGHGGRLARDQ 932
QY 721 CNLHRKPQELDFCVLRPC 738
Db 933 CNLHRKPQELDFCVLRPC 950

RESULT 2
Q91256 PRELIMINARY; PRT: 340 AA.
AC Q91256;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE Similar to a disintegrin and metalloproteinase with thrombospondin
DE motifs 1 (ADAMTS-1) (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC009667; AA09667.1; -.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 2.
KW Integrin.
FT NON_TER
SQ SEQUENCE 340 AA: 37021 MW; 06562C74763A8BD CRC64;

Query Match 41.7%; Score 1685; DB 11; Length 340;
Best Local Similarity 91.7%; Pred. No. 4.2e-147;
Matches 308; Conservative 11; Mismatches 17; Indels 0; Gaps 0;

QY 403 KCKLICRANGTGYFYVLAAPKVVVDGTLCPDSTSVCVQKCIKAGCDNGLSKRRFKDCGY 462
```

```
Db 5 KCKLICRANGTGYFYVLAAPKVVVDGTLCPDSTSVCVQKCIKAGCDNGLSKRRFKDCGY 64
QY 463 CGGDNKSKKVTGLFTKPMHGYNFVVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQGR 522
Db 65 CGGDNKSKKVTGLFTKPMHGYNFVVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQGR 124
QY 523 YLLNGHFVYVSAVERDLVYKGSLLRYSGTGTAVESLOASRPILPLAVEVLSVGMKTPPRV 582
Db 125 YLLNGHFVYVSAVERDLVYKGSLLRYSGTGTAVESLOASRPILPLAVEVLSVGMKTPPRV 184
QY 583 RYSFYLPKPREPREDKSSHPKDPGRGSPVLHNSVLSNQVEQDPPRPPARVWAGSWGPCSCAS 642
Db 185 RYSFYLPKPREPREDKSSHPKDPGRGSPVLHNSVLSNQVEQDPPRPPARVWAGSWGPCSCAS 244
QY 643 CGSGQKRAVDCRGSAQRTVPACDAARHVPETQACGECPTWELSAWSPCSCSGRGFQ 702
Db 245 CGSGQKRAVDCRGSAQRTVPACDAARHVPETQACGECPTWELSAWSPCSCSGRGFQ 304
QY 703 RSLKVCVGHGGRLARDQCCLRRKPFQELDFCVLRPC 738
Db 305 RSLKVCVGHGGRLARDQCCLRRKPFQELDFCVLRPC 340

RESULT 3
Q19791 PRELIMINARY; PRT: 2165 AA.
ID Q19791;
AC Q19791; Q27524;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE F25H8.3 protein.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Gajadasy S.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z69361; CAA93288.1; -.
DR EMBL; Z69360; CAA93288.1; JOINED.
DR EMBL; Z69360; CAA93287.1; -.
DR EMBL; Z69361; CAA93287.1; JOINED.
DR HSP; P15167; 1DTH.
DR MEROPS; M12.135; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_Mtpeptdse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 14.
DR SMART; SM00209; TSP1; 18.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSP1; 6.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 2165 AA; 244397 MW; FCC3DA8AAA9C4888 CRC64;
```

```
Query Match      38.8%; Score 1567; DB 5; Length 2165;
Best Local Similarity 39.6%; Pred. No. 4.8e-135;
Matches 321; Conservative 113; Mismatches 283; Indels 94; Gaps 17;

QY 7 YVELTWADSMVRFHGDLEHYLLTLATAARLYRHPSILNPNINIVVVKVLLLRDRDSG 66
D 281 YVELVWADTWMEYHGRSLEDYVTLFSIVASIRHQSLRASINIVVVKVLLVKTENAG 340
QY 67 PKVTGNAALTLRNCAQKLNKVKSDKHPEYDWTAILFTRODLCGAT--TCDTLGMA 125
D 341 PRITONAQOTLQDFCRQOQYNDPDDSSVQHHDAVAILTRKDCRSQKCDTLGLAELGT 400
QY 126 MCDPKRSCSVLEDDGLPSAFTTAHELGHVFNMPHDNKKVCEEV-----FGKL 172
D 401 MCDMQKSCAIIEDNGLSAAFTIAHELGHVFSIPHDDEKCKSTYPMNVKVKCFQSTKFDKT 460
QY 173 RAN---HMSPTLIQIDRANPWSACSAAIITDFLDSHG--DCLLDQPSKPISLPE--- 223
D 461 QFQNNFHIMAPTLLEYNTHPWSWSPCSAGMLERFLENNRGQTCLFQPVERRYYEDFVR 520
QY 224 DLPGASYLSQOELAFGVGSKPCPYQYCTKLWCTGKAKGMVQCOTRHFPWADGTSCGE 283
D 521 DEPGKKYDAHQCKFEVGPASLCPYMTCCRLWCATFYGSQMGCRTOHMPWADGTPCDE 580
QY 284 GK--LCUKGACVERHNLNKHVRDGSNAKWDYPGPCSRTCGGVQOLARRQCTNTPANGK 341
D 581 SRSMFCHGACVRLAPESLTKIDQWGDWRSWGECSRTCGGVQKGLRDCDSPKPRNGK 640
QY 342 YCEGVRYVYKSCNLEPCSSASGSKSFREEQCAFNGYN-----HSTNRLTLAVANVPKY 395
D 641 YCVQQRERYRSCNTQCEPMDT--QPYREVQCSEFNKNDIGIQVASTN-----THWVPKY 693
QY 396 SGVSPRDKCLICRANGTGYFYVLAIPKVVDGTLCSPDSTSVCVQKCIKAGCDGNLGRK 455
D 694 ANVAPNERCKLYCRLSGSAFYLLRDKVVDGTPCDRNGDDICVAGACMPAGCDHQLHSTL 753
QY 456 RFDKCGVCGGDNKSKKVTGLTKP--MHGVNFVVAIPAGASSIDIRQGYKGLIGDDNYL 514
D 754 RRDKCGVCGGDDSKKVKGTFTNEQGTFGYNEVMKIPAGSANIDIRQGYNNMKEDDNYL 813
QY 515 ALKNSOGKYLNLGHFVYSAVERDLVVKGLLRYSGTGTAVESLQASRPITLPTVEVLSV 574
D 814 SLRAANGEFLNGHFQVSLARQOIAFDVLEYSGDAIIERNGTPTRSDIYVHVLVS 873
QY 575 GKMTPTPRVYFSLPKPREP-----DKSS 598
D 874 GS-HPPDISYEMTAAPNAVIRPISSALYLWRTVDTWTECDRACRQOQSKMLCLDMST 932
QY 599 HPKDRPGPSVLHN-----SVLSLSNQVEQPPDRPPARWVAGSWGPCSSACSGLOKRAVDC 654
D 933 HRQS-----HNRNQNVLKPKQATRMCDICSTRWITEDYSSSACGSGKQRQVSC 985
QY 655 RGSAGORTVPA-----CDAAHRPVETQACGEP--TWELSAWSPCSKSG--RGFORSLK 707
D 986 VKMEGDRQTPASEHLCDRNSKFSDIASCYIDCGRKNWYGEWTSCTSEICGSGNKMHRKSY 1045
QY 708 CYCHGGRILLARQCNLHRPQELDFCVLRPC 738
D 1046 CVDDSNRRYVDESICGREGQKATERECNRIPC 1076

RESULT 4
Q8SXB0
ID Q8SXB0 PRELIMINARY; PRT; 1688 AA.
AC Q8SXB0;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DE GH1639p.
GN CG6107.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
```

[illegible]

Db 637 SQKCPRDS--VDPRACQAEHNSRRFRGRHYK-----WKPYQTQVEDODLCKLCYCI 684
 QY 410 ANGTGYFYVLAPKVVDGTLCSDPDSTSVQVQKCIKAGCDNGLSKKRFKDCGCGGDNKS 469
 Db 685 AEGFDFFSLSNKVKDTCPSDSRNVCDIGCICRVGCDNVGLGSDAEDVCGVCGNNSA 744
 QY 470 CKKVTGLTKPMH--GYNFWVAIPAGASSIDIRQRYGKLGIDDDNYLALKNSQKGYLLNG 527
 Db 745 CTIHRGLYTKHHTNOYHYHMVTPSGARSIRIYEMNV-----STSYISVRNALRYLNG 799
 QY 528 HFVSAVERDLVVKSLGSLRYSGTGAVESLQASRPILFPLTEVLSVGKMTTPRVRYSFY 587
 Db 800 HMTVDWPGR-YKFSGTYTDFRYSYNEPENLIATGPTNETLIVELLFQGR--NPGVAWEYS 856
 QY 588 LPKEPREDRKSHPKDPRGPS-----VLHNSVLSLSNOVE---QP 623
 Db 857 MPR-----LGTEKQPPAOPSYTWAIVSECSVSCGGQMTVREGCYRDLKQVNMFSFNP 911
 QY 624 DRDP-----PARWAGSWGPGCSASCGSLQKRAVDC--RGSAGQRTVPA--CD 667
 Db 912 KTRPVTGLVPCVKVSAACPPSPWSVGNNSACSRCTCGGAQSRPVQCTRRVHYDSEPVFASLCP 971
 QY 668 AARHPVETQAC--GEP--PTWELSAWSPCKSCGRGFQFBSRLSKCVGHG-----GRLLRQOC 721
 Db 972 QP-AFSSRQACNSQSCPPAWSAAGPNAECSTCGKWRKRAVACKSTNPASRAQLLPDAVC 1030
 QY 722 NLHRKPQELDFCVLRPC 738
 Db 1031 TSEPMPRMEACLLQRC 1047

RESULT 7
 Q9VF61 ID Q9VF61 PRELIMINARY: PRT: 1229 AA.
 AC Q9VF61:
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE CG6107 protein.
 GN CG6107.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler J., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matti B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003709; AAF55199.1; --
 DR FlyBase; FBgn0038340; CG6107.
 DR InterPro; IPR002870; Pep_M12B_propep.
 DR InterPro; IPR001590; Reprolysin.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR000130; Zn_MTpeptdse.
 DR Pfam; PF01562; Pep_M12B_propep; 1.
 DR Pfam; PF01421; Reprolysin; 1.
 DR Pfam; PF00090; tsp_1; 2.
 DR SMART; SM00209; TSP1; 3.
 DR PROSITE; PS00215; ADAM_MEPRO; 2.
 DR PROSITE; PS00192; TSP1; 2.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Zinc.
 SQ SEQUENCE 1229 AA; 138136 MW; C1293180AB19B700 CRC64;

Query Match 27.7%; Score 1119.5; DB 5; Length 1229;
 Best Local Similarity 34.5%; Pred. No. 4.9e-94;
 Matches 281; Conservative 111; Mismatches 260; Indels 163; Gaps 35;

QY 8 VETLVADESXVPHGADLEHLLTLATAARLYRHPESILNPINIVVKKVLLLRDRDSGP 67
 Db 356 LEVLIAVDNSMKQFHGDLQPIYLLILMSIVSSIFADASIGNSIRILLVRLISL-----P 409
 QY 68 KV---TGNAAALTLANCAWQKLNKVSXKHPEYWDTAIFTRQDLGAT---TCDILGMA 121
 Db 410 NINDQTHSSNEMLKHFQCF---INQGYER---DTAMLITREPICGSPVGKICHMLGLA 462
 QY 122 DVGTMCDPKRSVIEDDGLPSAFTTAHELCHVFENPHDNVKYCEEVFGKLRAN---HMM 178
 Db 463 ELGTVCSSSSCSIVQDTGLPTAFTMAHELGHILNNHDDDDKCMFYVTRQNNKVLHIM 521
 QY 179 SPLT-IQIDRANPWSACSAIITFDLSDHGDCLLDPQSKPISLP---EDLPASGYTLQ 234
 Db 522 SSVNGIHM---HPWSWSKK---TD-----KSCL--ETSVGAHIPYGTERTLPGEIYSLDA 567
 QY 235 QCELATGCGSKPCPYMOCYCTKLWC---TGKAKQMVQCTRHFFWADGTSCG-BGKLCILKG 290
 Db 568 QCQLSGNDRGYCTPTDEECKRLMCNRTSGNSNEQ---CASNLPWADGTPCGSSGHWCRG 625
 QY 291 ACVERHNLNKH---RVDSGSAWKDPYGCSTRTCGGQVQLARRQCTNPPTANGKCYCEGV 346
 Db 626 KCVS---NKHGYGRVQNGWGWPFTPTCSLTCGGVQESRECNQVPENGKKYCTGS 681
 QY 347 RVKYRSCNLEPCPSSASGSKFRQEQCEAFNGYHNHNRILTAVAWPKYSGVSPDRDKKL 406
 Db 682 RKRYRSCNTHQCPGSGNDP--REQQYAMNGRMNIPGVNPDTKWPKYE---KDACKL 735
 QY 407 ICRANGTGYFVLAPKVDGTLCSDPDSTSVQVQKCIKAGCDNGLSKKRFKDCGCGCD 466
 Db 736 FCRMDMKYVFMKLSMVTGDTSCAVDSFKCVNGICRPGACDNELMSIAKLDKCGVCEGR 795
 QY 467 NKSCKKVTG-LFTKPMHGYN-----FVVAIPAGASSIDIRQRYGKLGIDDDNYLA 515
 Db 796 NDTCHVETGNLLSNLLGLNDGNPNKTLVYVTRIPKASNIITQGYP---DQNFIV 851
 QY 516 LKNSQCKYLLNGHFVYSAVERDLVVKSLGSLRYSGTGAVESLQA---SRPLPLPTEVLVS 573
 Db 852 LTDDRDNELLNGKF-LKTYPLKPFYAGVTMQYTGSSSVVEQVNTTYSWLSRDLIVQIIS 910

QY 574 VGKMTTPR-----VRYSYFLPKBPPE-----594
D 911 LDVSPSKRQDTLLSYSTYIDKPDYAEVEIYRWEMQAPSNCDLSLCEGRSHRLPACIST 970
QY 595 -----DKSHPK-DPRGPSV---LHNSVLSLNSQVEQDDPPPPARWAGSWGPC 639
D 971 TQGVKVAPOFCDSAMPKIDDRACNTDCLRLNLAVTSTSE-----C 1010
QY 640 SASGCS-GLQKRAVDC---RSAGORT---VPACDAAHRPVEVTAQGECPCTWELSAWSP 692
D 1011 SAACGELGTREKTVACVQTTNNQORSNIVDMSTCKLKFVAYHEECREGC--WVLSWST 1068
QY 693 CSKSCGRGFRQRSLKCVGHGRLRLAQDQCNLHRKP 727
D 1069 CSKSCGTGSGQREAHCHYLNHSR-VSDCLCNPRTKP 1102
RESULT 8
Q9W493
ID Q9W493 PRELIMINARY; PRT: 1054 AA.
AC Q9W493;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CG4096 protein.
GN CG4096.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Culey S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M.C., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT *The genome sequence of Drosophila melanogaster.*;

RL Science 287:2185-2195(2000).
DR EMBL; AE003435; AAF46065.1; --
DR HSP; P15167; IATL.
DR MEROPS; M12.231; --
DR FlyBase; FBgn0029791; CG4096.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Repolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_Mrpeptdse.
DR Pfam; PF015621; Pep_M12B_propep; 1.
DR Pfam; PF00090; tsp-1; 2.
DR SMART; SM00209; TSPI; 2.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSPI; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 1054 AA; 118616 MW; DC15455555CB6212 CRG64;
Query Match 27.6%; Score 1117; DB 5; Length 1054;
Best Local Similarity 35.0%; Pred. No. 6.6e-94;
Matches 276; Conservative 116; Mismatches 276; Indels 120; Gaps 34;
QY 2 VSIPRYVETLVVAVDESMVKFHGADLEHYLLTLATAARLYRHPSPINPINIVVVKVLLLR 61
D 321 ISSPRVETLIVADATMSAFH-RDLNGYLLTIMNVVSALYKDPSSIGNSIIVVRIQLD 379
QY 62 DRDSGPK--VTGNAALTLRNFCAMOKKLNKVSUKHPEYWDFTAILTRQDLGATTCDTLG 119
D 380 EESQLQLNTAQNALDRFCQWHLKMGSEKDPHHHDVAILLITRKNIC--ANNCMTLG 438
QY 120 MADVGTMCDPKSCSVIEDDGLPSAFTTAHGLGHVFNPHDNVKV-CEEVFCKLRANMM 178
D 439 LANVGCMCKPKSCSVNEDNGIMLSHTITHLGHNFMGPHDTAKIGCHPRVGPV--VHIM 496
QY 179 SPTL----TIDRANPWSACSAAITDFLDGSHGDCLLDQPSKPI---SLPEDLPASVT 231
D 497 TPTFGADITLV---CWSNCSRKTYTHFLDQGLGEC-LDDPTPLDEYNTGELPGMYN 551
QY 232 LSQOCELAFG-----VSGKPCPYMOYCTKLMCTKAKGMVCQTRHFFWADGTCGEGK 285
D 552 ARGQCRQLFNLTATDSEVGACSAFH-EFCSTLWC--KVNGE--CVTHMRPTAGTLCGRNK 606
QY 286 LCLKGACVSRHNLKRVDSWAKWDPYPCSRCTGGVGLARROCTNTPANGKCYCBG 345
D 607 WCQNGKCVREEL--AAVNGGWDSEWSECSRGSGVSTOQRECDNPVANGVFCIG 664
QY 346 VRVYRSCNLEPCSSASGSKSFEEQCEAFN--GYNHSTNRLTLAVAWVPKYSGVSPDK 403
D 665 ERKRYKICRKPCCP--AEPSPFAQCACAFDNVSYQATYK-----WLPFFDKNPP-- 713
QY 404 CKLICRANGTGYVVLAPKVVDTGLCSPTSTVCVCGKCIKAGCDGNLGSKKRFDKCYC 463
D 714 CKLFCSDVDDTITANNGATVLDGTCTGTNNNCIDGICKKVGCDWIDVSEVDDRCGVC 773
QY 464 GGDNCKCKVTGFTKPM-----HGYNFVVAIPAGASSIDIRORYKGLIGDNNYLAKNS 519
D 774 GGSQDQCPVRETYTDFPAKDGAYEIVTIPARAHILIRE-----LANSFHLAATG 828
QY 520 QG--KYLINGHFVYSAVERDLVVKSLRYSCTGTAVESLQASRPILPLETVEVLSVGM 577
D 829 DGGDRFYNGDSLIS-MPGFEIAGAESLYDRVDEQ-ETITTIPOIQTQHSISLYAIVRGNE 886
QY 578 TPRVRYSYFLPREPKEDKSSHPKDPGRPSVLSNLSNQVEQDDPPPPARWAGSWG 637
D 887 SNAGIFYEFTLP-----ALN---VTAGROFQ-----WRLSNWT 916
QY 638 PCSASCSGSLQKRAVDC---RGSAGO-----RTVPACDAA--HRPV- 673
D 917 ACSASCGGVQVHREPCQENGKESNEPQRIYSIKRNLIIALGDTLPCWTHAKNRPAR 976
QY 674 ETQACQ-EPQPT-WEISAMSPCSKSGR-GF-----QRRSLKCVGHGRLRLAQDQCNLHR 725
D 977 QSRGCGQPCPAHWPQPMQFCVTPRPVGFAPQRRRSVSVCLDEHVDVVADEACGHLQ 1036


```
QY 726 KPQELDFC 733
Db 1037 KPAEMEPC 1044

RESULT 9
Q8TE56 PRELIMINARY; PRT; 1095 AA.
AC Q8TE56;
DT 01-JUN-2002 (TREMELrel. 21, Created)
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Metalloprotease disintegrin 17, with thrombospondin domains.
GN ADAMTS17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

SEQUENCE FROM N.A.
RP MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ315735; CAC86016.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1095 AA; 121099 MW; A5C6540484754D5D CRC64;

Query Match 26.1%; Score 1057; DB 4; Length 1095;
Best Local Similarity 32.6%; Pred. No. 2.5e-86;
Matches 261; Conservative 109; Mismatches 290; Indels 140; Gaps 27;

QY 8 VETLVVADSMVKFHGAD-LEHYLLTLLATAARLYRHPHSILNPINIVVKKVLLLRDRSG 66
Db 234 VETLVVADSMVKFHGAD-LEHYLLTLLATAARLYRHPHSILNPINIVVKKVLLLRDRSG 66
QY 67 PKVTGNAALTLRNFCAWQK-----LNKV--SDKHPEYDTAILFTRODLC--GAT 113
Db 294 LSIHHGERSLESCHWQNEEYGGARYLGNQVPGGKDDPLVDAAVFTKDFCVHKDE 353
QY 114 TCDILGMADVTCMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKKVCEEVFGKLR 173
Db 354 PCDVTGIAYLGGVCSAKRKCVLAEENGLNLAFTIAHELGHNLGMND-----DDHSSCAG 408
QY 174 ANHMSPTLIQIDRANP---WSACSAAILTDFLDLSDGDCLL---DQSPKPTSLPEDLP 226
Db 409 RSHINSGEWK--GRNPSDLWSWSSCRDDEFLKSKVSTCLLVTPRSQHTVRLPHKLP 466
QY 227 GASTYLSQOCBELAFGVGKPCPYMQY--CTKLWCTGKAKGQMVQCTRHFFWADTSCGEG 284
Db 467 GMHYSANEQCOILFGMNATFCRNHEHLMCAGLWCL--VEGDTCKTKLDPDTECGAD 524
QY 285 KLCLKGACVERHNLKRVDSWAKWDPYGPCSTCGGVQLARRQCTNTPANGKYCE 344
Db 525 KWCRAECVSTPIPEH--VDGDWSPGAWSMCSRTCTGCGARFRQKCDNPPPGPGTHCP 583
QY 345 GVRKYRSCNLEPCPSSASGSKFRFEEQCEAFNGYNSTNRLTLAVANPKYSGVSPDKC 404
Db 584 GASVEHAVCNLPCPKGL--PSFDQCOQAHDRLSPKKGLLTV-----VVDKPC 633
QY 405 KLICRANGTGFFYVLAIPKVDGTLCSPDSTSVCVQGVKICIKAGCDNGLSKRFDKCGVC 464
Db 634 ELYCSPLGKESPLLVADRVLDGTPCGPYETDLCVHGKQKIGCDGIIIGSAKEDRCGVCS 693
QY 465 GDNKSKVTGLFTKPMHGVNFVVAIPAGASSIDIRGFKGLIGDNDYALKNNSQKYL 524
Db 694 GDGATCHLVKGDGFSH-----ARGTALKDSG--KGSI-----NSDWKIE 729
QY 525 LNHGFVVAVERDLVWKGSLRLRYSGTGATVESLQASRPILPTVEVL-----SVGKMT 579

Db 730 LPGEFOIA-----GTTVRYVRRG-LWEKISAKGPTKLPLHLMLLFLHFDQDYG---- 775
QY 580 PRVRYSEYLPKPREDKSSHPKDP-----GPSV-----LHNS 612
Db 776 --IHVEYTVPNRTAENQSEPEKPDLSLFIWTHSGWEGSCVQCGGGRRTIVSCTRIVNK 833
QY 613 VLSLSNOVEQDDRRPP-----ARWAGSWGPCSCASGSGLOKRAYDC-----R 655
Db 834 TTTLVNDSDCPQASRPEFQVRRRCNLHPCQSRWVAGWSPCSATCEKGFQHREVTCVYQLQ 893
QY 656 GSAGQRTVPACDAARHPEVETQAC-GEPC-PTWELSAWSPCSKSGRGFORSLKCVHGG 713
Db 894 NGTHVATRPILCYCPGPRPAAVQSCGQDCLSIWEASEWSQCSASCGKGVWKTIVACTNSQ 953
QY 714 RLZARDQCNLHRKPOELDFC 733
Db 954 K-----CDASTRPRAEAC 967

RESULT 10
Q8WX58 PRELIMINARY; PRT; 1223 AA.
AC Q8WX58;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE A disintegrin-like and metalloprotease with thrombospondin type 1
DE motif 14 precursor.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

SEQUENCE FROM N.A.
RP MEDLINE=21638061; PubMed=11779638;
RA Bolz H., Ramirez A., von Brederlow B., Kubisch C.;
RT "Characterization of ADAMTS14, a novel member of the ADAMTS
RT metalloproteinase family.";
RL Biochim. Biophys. Acta 1522:221-225(2001).
DR EMBL; AF358666; AAL40229.1; -.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; TSP-1; 4.
DR SMART; SM00209; TSP1; 4.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS0092; TSP1; 1.
KW Signal; Integrin; Protease; Metalloprotease.
FT SIGNAL 1 22
SQ SEQUENCE 1223 AA; 133871 MW; 3639AC4D92F170F CRC64;

Query Match 25.3%; Score 1024; DB 4; Length 1223;
Best Local Similarity 31.7%; Pred. No. 3.2e-85;
Matches 265; Conservative 95; Mismatches 296; Indels 180; Gaps 32;

QY 8 VETLVVADSMVKFHGAD-LEHYLLTLLATAARLYRHPHSILNPINIVVKKVLLLRDRSG 66
Db 261 IEVLLVDDSVVRHFQKEHVQNVLTLMIVDEIYHDESIGLHINIALVRLIMVGYRQL 320
QY 67 PKV-TGNAALTILRNFCAWQKLNKVSDEKHPYWDTAILFTRODLCGATCTDLGMADYGT 125
Db 321 SLIERGNPSRLEQVCRWAHSQORQDPQSHAEHHDHVFLTRQDF-----GPSGYAPVTG 374
QY 126 MCDPKRCSVIEDDGLPSAFTTAHELGHVFNPHD---NVKVCVEVFGKLRANHMSPTL 182
Db 375 MCHPLRSCALNHEDGFSFAFVIAHETGHVGLMEHDGOGNGCADETSLGS-----VMAFLV 429
QY 183 IQIDRANPNSACSAAILITDFLDGSHGDCULLDQSPKPT-SLPEDLPGASYTSLSQCELAG 241
```


QY	649	KRAVDCRGAGQRTVPACD-----AAHRPVTQAC-GEPCPT-WELSAWSPCSKSGRGFQ	702
Db	998	SRQVACTQQLSNGTLIRARERCDCIGPKPASAQCEGQDCMTVWEAGVWSEFSVKCKGKGR	1057
QY	703	RRSLKCVGHGRLARDOCNLHRKPOELDFC	733
Db	1058	HRTVRCIN-----PRKKCVLSRPREADC	1082
RESULT 14			
Q96L37	PRELIMINARY; PRT; 1427 AA.		
AC	Q96L37	01-DEC-2001 (Tremblrel. 19, Created)	
DT	01-DEC-2001	(Tremblrel. 19, Last sequence update)	
DT	01-JUN-2002	(Tremblrel. 21, Last annotation update)	
DE	Von Willebrand factor-cleaving protease precursor.		
GN	ADAMTS13.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=LIVER;		
RX	PubMed=11557746;		
RA	Zheng X., Chung D., Takayama T.K., Majerus E.M., Sadler J.E.,		
RA	Fujikawa K.;		
RT	"Structure of von Willebrand Factor-cleaving Protease (ADAMTS13), a		
RT	Metalloprotease Involved in Thrombotic Thrombocytopenic Purpura,"		
RL	J. Biol. Chem. 276:41059-41063(2001).		
DR	EMBL: AY055376; AAL17652.1;		
DR	MEROPS: M12.241;		
DR	InterPro: IPR001590; Reprolysin.		
DR	InterPro: IPR000884; TSPL.		
DR	InterPro: IPR00130; Zn_MTPeptdse.		
DR	Pfam: PF01421; Reprolysin; 1.		
DR	Pfam: PF00090; tsp.1; 4.		
DR	PROSITE: PS0215; ADAM_MEPRO; 1.		
DR	PROSITE: PS50092; TSPL; 1.		
DR	PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.		
KW	Protease; Signal.		
FT	SIGNAL		
FT	CHAIN		
FT	SEQUENCE		
QY	75	VON WILLEBRAND FACTOR-CLEAVING PROTEASE.	
QY	1427	AA: 153632 MW: 81BC3AABC1A4442 CRC64;	
Query Match 21.48; Score 865; DB 4; Length 1427;			
Best Local Similarity 29.6%; Pred. No. 2e-70;			
Matches 234; Conservative 103; Mismatches 328; Indels 126; Gaps 30;			
QY	7	VYETLVVADSMVKFPGADLEHLLTLTAARLYRHPSILNPINIVVVKVLLLRDRDSG	66
Db	81	HLELLVAVGPDVQAHQEDTERVLTNLNIGALLRPSLGAQFRVHLVKMVLTEPEGA	140
QY	67	PKVTGNAALTNRNCAWQKLNKVSQKHPEYWDTAILFTRQDL-CGATCTDLGMADVGT	125
Db	141	PNITANLTSSLLSVCGSQTNIPEDDTPGHADLVLYITREDFLELPDGNRQVRGVTOLGG	200
QY	126	MCDPRKRSVIEDGLPSAFTTAHELGHVENPHDNVKNVCEEVFGKLRANHMSPFTLIQI	185
Db	201	ACSPTWSCLTIEDTDFDGLVTAHGHSHFLEHGDGAPGS-----CGPSGHVMA-----	251
QY	186	DRANP-----WSACSAIITFDLSGHGDCLLD-----QPSKPISLPEDLPQASVTL	336
Db	252	DGAAPRAGLAWSPCSRQLLSLSAGARCVWDPRPQPGSAGHPDPAQPLGYTSANEQC	311
QY	237	ELAFGVGSKPCPY-----MQYCTKLWCTGKAGQWVCQTRHPFWADGTSCGEGKLC	392
Db	312	RVAFGPKAVACTFAREHLDMQALSCHTDPLDQSSCSRLLYPLLDGTEGVEKWC	371
QY	293	VERHNLNK-HRVDSMAKWDYPCSRCTCGGVQLARRQCNTNPTPANGKYCEGVVRKYR	351
Query Match 20.5%; Score 830; DB 6; Length 269;			
Best Local Similarity 54.7%; Pred. No. 3.1e-68;			
Matches 146; Conservative 37; Mismatches 74; Indels 10; Gaps 3;			

```
QY 117 TLGMADVTMCDPKRSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCCEEVFGKLRANH 176
Db 1 TLGMADVTCTCSPERSCAVIEDDGLHAFTVAHEIGHLGLSHDDSKFCEENFGSTEDKR 60
QY 177 MMSPTLIQIDRANPSACSAAIITDFLDSGHGDCLLDOPSKPISLPEDLPGASYTLSoQC 236
Db 61 LMSSILTSIDASKPSKCTSATITEFLDDGHGNCLLDVFRKOILGPEELPGQTYDATOOC 120
QY 237 ELAFVGSKPCPYMYCTKLWCTGKAKGQMVQOTRHFPWADGTCGEGKLCILKGACVER- 295
Db 121 NLTFGPEYTVCPGMDVCARLWCNAVVRQGMVCLTKKLPAVEGTPCGKGRICLQKCVDKT 180
QY 296 ----HNLNKHRYDGSWAKWDYPGCSRTCGGQVQLARQCTNPTPANGGKYCEGVVRKYR 351
Db 181 KKKYISTSSH---GNWGSWGPWGQCSRGCGGVQFAYRHNNPAPRNSGRYCTGKRATYR 237
QY 352 SCNLEPCPSSASGKSFREOCEAFNGY 378
Db 238 SCSVTPCP--ANGKSFREOCEAKNGY 262
```

Search completed: April 29, 2003, 17:18:02
Job time : 54.0027 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 9.08922 Seconds
(without alignments)
3367.676 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_950

Perfect score: 4043

Sequence: 1 FVSIPIRYVETLVVADESMVK.....DQCNLHRKPQLDFCVLRPC 738

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2257	55.8	968	1 ATSL_MOUSE	P97857 mus musculus
2	2254	55.8	967	1 ATSL_HUMAN	Q9uh18 homo sapien
3	2251	55.7	967	1 ATSL_RAT	Q9wug1 rattus norv
4	1911.5	47.3	905	1 ATSL_MOUSE	P57110 mus musculus
5	1900	47.0	890	1 ATSL_HUMAN	Q9up79 homo sapien
6	1779	44.0	1629	1 ATSL_HUMAN	Q9p2n4 homo sapien
7	1761	43.6	837	1 ATSL_HUMAN	O75173 homo sapien
8	1745	43.2	930	1 ATSL_HUMAN	Q9una0 homo sapien
9	1730.5	42.8	630	1 ATSL_RAT	Q9esp7 rattus norv
10	1728	42.7	930	1 ATSL_MOUSE	Q9r001 mus musculus
11	1184.5	29.3	1593	1 ATSL_HUMAN	P58397 homo sapien
12	1155.5	28.6	1077	1 ATSL_HUMAN	Q9h324 homo sapien
13	1094	27.1	997	1 ATSL_HUMAN	Q9ukp4 homo sapien
14	1075	26.6	1211	1 ATSL_HUMAN	O95450 h adamts-2
15	1062.5	26.3	1205	1 ATSL_HUMAN	O15072 homo sapien
16	1045.5	25.9	1205	1 ATSL_BOVIN	P79331 b adamts-2
17	942.5	23.3	860	1 ATSL_HUMAN	Q9ukp5 homo sapien
18	601.5	14.9	245	1 ATSL_BOVIN	Q9tt93 bos taurus
19	599	14.8	207	1 ATSL_BOVIN	Q9tt92 bos taurus
20	381.5	9.4	450	1 ATSL_MOUSE	P58459 mus musculus
21	267.5	6.6	956	1 AD19_HUMAN	Q9h013 homo sapien
22	260	6.4	776	1 AD28_MACFA	Q9xsl6 macaca fasc
23	258.5	6.4	571	1 DISU_BOTJA	P30431 bothrops ja
24	255.5	6.3	1584	1 BAIL_HUMAN	O14514 homo sapien
25	253.5	6.3	1074	1 SMSA_HUMAN	Q113591 homo sapien
26	253	6.3	824	1 AD08_HUMAN	P78325 homo sapien
27	250.5	6.2	775	1 AD28_HUMAN	Q9ukg2 homo sapien
28	248.5	6.1	1170	1 TSP1_HUMAN	P07996 homo sapien
29	248	6.1	1077	1 SMSA_MOUSE	Q62217 mus musculus
30	247.5	6.1	1093	1 SMSB_MOUSE	Q60519 mus musculus
31	245.5	6.1	1170	1 TSP1_MOUSE	P35441 mus musculus
32	244.5	6.0	920	1 AD19_MOUSE	O35674 mus musculus
33	244.5	6.0	1170	1 TSP1_BOVIN	Q28178 bos taurus

34	238	5.9	774	1 AD28_MOUSE	Q9jln6 mus musculus
35	237.5	5.9	813	1 AD33_HUMAN	Q9bz11 homo sapien
36	236.5	5.8	909	1 AD12_HUMAN	O43184 homo sapien
37	233.5	5.8	903	1 AD12_MOUSE	Q61824 mus musculus
38	221	5.5	857	1 AD22_MOUSE	Q9rlv6 mus musculus
39	220.5	5.5	906	1 AD22_HUMAN	Q9p0k1 homo sapien
40	220	5.4	416	1 HRIB_TRIFL	P20164 trimeresuru
41	219	5.4	814	1 AD15_HUMAN	Q13444 homo sapien
42	219	5.4	826	1 AD08_MOUSE	Q05910 mus musculus
43	208.5	5.2	1172	1 TSP2_MOUSE	Q03350 mus musculus
44	208.5	5.2	1173	1 TSP1_XENLA	P35448 xenopus lae
45	207.5	5.1	797	1 AD33_MOUSE	Q923w9 mus musculus

ALIGNMENTS

RESULT 1
ATSL_MOUSE
ID AC P97857; OS4768; STANDARD; PRT; 968 AA.
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (SC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN ADAMTS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=98110583; PubMed=9441751;
RA Kuno K., Lizasa H., Ohno S., Matsushima K.;
RT "The exon/intron organization and chromosomal mapping of the mouse
RT ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";
RL Genomics 46:466-471(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97150761; PubMed=8995297;
RA Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.,
RA Matsushima K.;
RT "Molecular cloning of a gene encoding a new type of metalloproteinase-
RT disintegrin family protein with thrombospondin motifs as an
RT inflammation associated gene.";
RL J. Biol. Chem. 272:556-562(1997).
RN [3]
RP CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.
RX MEDLINE=99303657; PubMed=10373500;
RA Kuno K., Terashima Y., Matsushima K.;
RT "ADAMTS-1 is an active metalloproteinase associated with the
RT extracellular matrix.";
RL J. Biol. Chem. 274:18821-18826(1999).
RN [4]
RP FUNCTION.
RX MEDLINE=20389568; PubMed=10930576;
RA Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.,
RA Ohno H., Matsushima K.;
RT "ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";
RL FEBS Lett. 478:241-245(2000).
RN [5]
RP FUNCTION, AND INDUCTION.
RX MEDLINE=20243757; PubMed=10781075;
RA Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W.,
RA Richards J.S.;
RT "Progesterone-regulated genes in the ovulation process: ADAMTS-1 and
RT cathepsin L proteases.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).
CC -!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC SIMILARITY). ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH

CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1691-GLU-1-LEU-1692
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX.
CC -1- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY
CC INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY
CC LIPOPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA
CC CELLS OF PREOVULATORY FOLLICLES.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 7.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB001735; BAA24501.1; ALT_INIT.
CC EMBL; D67076; BAA11088.1; ALT_FRAME.
CC MEROPS: M12.222;
CC MGD; MGI:109249; Admts1.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR02870; Pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; Zn_MTpeptidase.
CC Pfam; PF00090; tsp.1; 3.
CC Pfam; PF01421; Reprolysin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC SMART; SM00209; TSP1; 3.
CC PROSITE; PS00215; ADAM_MEPRO; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE; PS50092; TSP1; 3.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolase: Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 48
FT PROPEP 49 253
FT CHAIN 254 968
FT SITE 206 206
FT METAL 402 402
FT ACT_SITE 403 403
FT METAL 406 406
FT DOMAIN 477 559
FT DOMAIN 560 617
FT DOMAIN 618 725
FT DOMAIN 726 850
FT DOMAIN 851 909
FT DOMAIN 910 968
FT DOMAIN 195 199
FT CARBOHYD 548 548
FT CARBOHYD 721 721
FT CARBOHYD 765 765
FT CARBOHYD 783 783
FT CARBOHYD 946 946
FT MUTAGEN 403 403
FT CONFLICT 335 335
FT CONFLICT 425 425
FT SEQUENCE 968 AA; 105841 MW; 42EBDA55499FB6C1 CRC64;

Query Match 55.8%; Score 2257; DB 1; Length 968;
Best Local Similarity 53.6%; Pred. No. 1.2e-156;
Matches 400; Conservative 130; Mismatches 176; Indels 40; Gaps 10;

QY 1 FVSPRVVETLVVADESVMKFGADLHYLLTLLTAARLYRHPHSILPINIVVVKVLL 60
DB 254 FVSPRVVETMLVADQSWADFGSLGKHYLLTFSVAARFYKHPHSIRNSISLVVVKILVI 313
QY 61 RDRSGPKVVTGNAALTFRNFCWQKLNKVKSDHPEYDAILFTRODLGSGHCTDLGM 120
DB 314 YEEQKGPVTSNAALTFRNFCWQKLNKVKSDHPEYDAILFTRODLGSGHCTDLGM 373
QY 121 ADVGTMCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 180
DB 374 ADVGTVCDPSRSCSVIEDDGLQAAFTTAHELGHVFNPHDDAKHSCASUNGVTGSHLMAS 433
QY 181 TLIDRANPWSACSAAIITDLDGSGHGDCLLDOPSPKISLPEDLPGLGASYTSLQOCELAF 240
DB 434 MLSSLDHSQPSPCSAIYVWTFELNGHGECLMDKPNKLPDLPGTLYDANRCQOFTF 493
QY 241 GYSGKPCP-YMQYCTKLMCTGKAKGQMVYCTRHPFPAWDTSCGEGKLCCLKACACVERHNLN 299
DB 494 GEESKHCPDAASTCTTLCWCTGTSGLLVCQTKHFPWADTSCGEGKVCVSKYCNKTDN- 552
QY 300 KH---RVDSNAKWDPYPCSRCTCGGQVLARRCTNTPANGGKYCRGVKVRSCNLE 356
DB 553 KHFAFVHSGMGPWGPWGDSCSTCGGVQYTMRECDNVPKNGGKYCRGVKVRSCNIE 612
QY 357 PCPSASCKSPRECEAFNGYNHSTNRLTLAVAMVPKYSGVSPRDKCLKICRANGTYGF 416
DB 613 DCPDN-NKGTFRFEOCEAHNEFSKASFGNEPTVETPKYAGVSPKDRCKLTCEAKGIGYF 671
QY 417 YVLAPKVVDTLCSPDSTSVCGQKICIRAGCDNGLSKRRDKCGVCGDKNKSKKVTGL 476
DB 672 FVLQPKVVDTLCSPDSTSVCGQKICIRAGCDRIIDSKKFKDCGVCNGNSTCKKMSGI 731
QY 477 FTKPHGYNFVVAIPAGASSIDIRORYKGLIGDDNYLAKNSOCKYLLNGHGFVVSAYER 536
DB 732 VTSTPRGYHDIVTIIPAGATNIEVKRNONGSRNNGSFIAIRADDTYILNGNFTLSTLEQ 791
QY 537 DLVWKGSLRLYSGTGTAVESLQASRPILPTEVLSVSGKMTPPRRVYSFYLPKPRDCK 596
DB 792 DLTGKTVLRYSSSAALERISFSPLEPLTIQVLWGHALRPKIKFTYFMKKKTES-- 849
QY 597 SHPKDPRGSPVLHNSVLSLSNOVQPDPRPARVAGSWGSCSAGSLQKRAVDCRG 656
DB 850 -----FNAIPTFS-----EWVIEWEGCSKTCGSGWRRVVQCRD 884
QY 657 SAGORTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCSKSGRGFQRRSLKCVGHGG 713
DB 885 INGH---PASECAKEVKVPASTRPCADLPCPHWQVGDWSPCSKTCGKYKRTLKCVSHDG 941
QY 714 RLARDQCNLHRRKQPE-LDFCVLRPC 738
DB 942 GVLNSCSDPLAKPKHYIDFCTLTQC 967

RESULT 2
ATSL_HUMAN STANDARD; PRT: 967 AA.
ID ATSL_HUMAN Q9UHP80; Q9UHP83; Q9P2K0; Q9NSJ8;
AC Q9UHP80; Q9UHP83; Q9P2K0; Q9NSJ8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).
GN ADAMTS1 OR METH1 OR KIAA1346.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.


```
Db 215 FVSEARFVETLLVADASMAAFYGADLQNHILTLMSVAARIYKHPISIKNSINLMVVKVLIV 274
.OY 61 RORDSGPKVTGNAALTNRNFCWQKLNKYSKDEYWDTAILFRQDLGAT-TCDPLG 119
Db 275 EBEKMGPEVSDNGGLTLNFRNFWQRFRNQPDSRHPHEHYDTAILFRQNFQCGEGLCDPLG 334
.OY 120 MADVGTMCDPKRSVIEDGLPSAFTTAHELGHVFNPHDNVKNVCEVEFGKLRANHMS 179
Db 335 VADIGTICDPNKSVCVIEDGELQAHTLAHELGHVLSLPHDSDCKTFLFPGMKHHVNA 394
.OY 180 PTLIQIDRANPWSAAIITDFLSGHGDCLLDQSPKISLPEDLPG--ASYTILSQOCE 237
Db 395 PLFVHLNQTLPWSPCSAMYLTELLDGHGDCLLDAPGAALPLTGLPGRMALYOLDQOCR 454
.OY 238 LAFGVGSKPCP---YMQVCTKLWCTGKAKGOMVCTRH--FPWADGTSCGEGKCLKAG 292
Db 455 QIFGPDFRHCPTNSAQDVCAQWCHTDG-AEPLCHTKNGSLUPWADGTPCGPHLCSEGC 513
.OY 293 VERHNLNKHRR--VDGSAWKPDPYCSRTCGGQVGLARRQCTNPTTPANGKYKVCGRVKY 350
Db 514 LPDEEVEVRPKPVVDGWAHPGWGBCSRTCGGQVGFSHRECKDPQPONGRYCLGRRAKY 573
.OY 351 RSCNLEPCPSSASGSRFEEQCEAPNGYHNSTNRLTLAVAVPKYSGVSPRDKCKLICRA 410
Db 574 QSCHEECP--PDGKSRFEEQCEKYNNYTDMDGNL-LQWVPKYAGVSPRDKCKLCFRA 630
.OY 411 NGTGYFYVLAPKPVVDGTLCSPDSTSVCGQKICRAGCDGNLGSKRFDKCGVCGDNKSC 470
Db 631 RGRSEKFEAKVIDGTLGCPETALCVRGOCVRAGCDHVVDSPRKLKDCGCVCGGNKSC 690
.OY 471 KVTGLGTPKMHGVNFVAIPAGASSIDIRQYKGLIGDDNYLALKNQSKQYLLNGHFV 530
Db 691 RYVSGSLTPTNYGNDIVTIPAGATNIDVKQSRHPGVQDNGNYLALKTADQYLLNGNLA 750
.OY 531 VSAVERDLVWGLSLRYSGTAVESLOASRPILLEPLTVEVLSV-GKMTPPRVRYSYLP 589
Db 751 ISAIQDILVGLTKLYSGSIATLRLQSFRLPEPLTVQLLTPVGEVFPKRYTTFVP 810
.OY 590 KEPRDKSHPKDPRGSPVHNSVLSLSNQVEQDPRRPARWAGSWGPCSASCSGLQK 649
Db 811 ND--VDFSMQSSKERATTNIIQLLH-----AQWVLGDSECSSTCGAGQWR 855
.OY 650 RAVDCRGAGORTVPACDAARHPVETQAC-CEPCP 683
Db 856 RTVECRDPGQASA-TCKALKPEDAKPCESQLCP 889
RESULT 6
IDS9_HUMAN
AC Q9P2N4; Q9NR29; STANDARD; PRT; 1629 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).
GN ADAMTS9 OR KIAA1312.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE=Fetal;
RX MEDLINE=20396138; PubMed=10936055;
RA Clark M.E., Kerner G.S., Turbeville L.A., Boyer A., Arden K.A.,
RA Maki R.A.;
RT "ADAMTS 9, a novel member of the ADAM-TS/Metallospandin gene
PL family",
PL Genomics 67:343-350(2000).
RN [2]
RN SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;
```

```
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RL for large proteins in vitro.",
RL DNA Res. 7:65-73(2000).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES.
CC EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG,
CC PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN
CC COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR
CC THYMUS.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF261918; AAF89106.1; --
DR EMBL; AB037733; BAA92550.1; --
DR HSSP; P15167; IATL.
DR MEROPS; M12.021.
DR Genew; HGNC:13202; ADAMTS9.
DR MTM; 605421;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTPeptidase.
DR Pfam; PF00090; tsp_1; 11.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPl; 12.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS00092; TSPl; 9.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Alternative splicing.
FT SIGNAL; 1 18
FT PROPEP; 19 287
FT CHAIN; 288 1629
FT DOMAIN; 509 587
FT DOMAIN; 589 642
FT DOMAIN; 645 752
FT DOMAIN; 753 880
FT DOMAIN; 999 1053
FT DOMAIN; 1056 1108
FT DOMAIN; 1111 1156
FT DOMAIN; 1184 1239
FT DOMAIN; 1240 1295
FT DOMAIN; 1332 1383
FT DOMAIN; 1445 1498
FT DOMAIN; 1501 1554
FT DOMAIN; 1562 1612
FT DOMAIN; 88 96
FT SITE; 223 223
CYSTEINE SWITCH (POTENTIAL).
```


RX MEDLINE-99395124; PubMed=10464288;
 RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
 RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
 zinc metalloproteases.";
 RL J. Biol. Chem. 274:25555-25563(1999).
 CC -I- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
 CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
 CC DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
 CC PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
 CC -I- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-1-Ala-393
 CC site.
 CC -I- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -I- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (by similarity).
 CC -I- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA
 CC BUT ALSO IN OTHER TISSUES, SUCH AS HEART AND BRAIN, AND ALSO
 CC CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE,
 CC CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN
 CC ARTHRITIC PATIENT.
 CC -I- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -I- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -I- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -I- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF142099; AAD49577.1; -;
 CC EMBL; AP001698; BAA95504.1; -;
 CC EMBL; AP001697; BAA95503.1; -;
 CC EMBL; AF141293; AAF02493.1; -;
 CC HSP; Q9PW35; 1BDU;
 CC MEROPS; M12.225; -;
 CC Genew; HGNC:221; ADAMTS5.
 CC MIM; 605007; -;
 CC InterPro; IPR001762; Disintegrin.
 CC InterPro; IPR002870; Pep_M12B_propep.
 CC InterPro; IPR001590; Reprolysin.
 CC InterPro; IPR000884; TSPL.
 CC InterPro; IPR000130; Zn_M12B_propep.
 CC Pfam; PF00090; tsp_1; 2;
 CC Pfam; PF01421; Reprolysin; 1.
 CC Pfam; PF01562; Pep_M12B_propep; 1.
 CC SMART; SM00209; TSPL; 2;
 CC PROSITE; PS0215; ADAM_MEPPO; 1.
 CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 CC PROSITE; PS50092; TSPL; 1.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Extracellular matrix.
 FT SIGNAL 1 16
 FT PROPEP 17 261 POTENTIAL.
 FT CHAIN 262 930 ADAMTS-5.
 FT SITE 209 209 CYSTEINE SWITCH (POTENTIAL).
 FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 411 411 BY SIMILARITY.
 FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 485 566 DISINTEGRIN-LIKE.
 FT DOMAIN 567 623 TSP TYPE-1 1.
 FT DOMAIN 624 731 CIS-RICH.
 FT DOMAIN 732 874 SPACER.
 FT DOMAIN 875 930 TSP TYPE-1 2.
 FT DOMAIN 37 41 POLY-ALA.
 FT DOMAIN 257 261 POLY-ARG.

FT	CARBOHYD	498	498	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	728	728	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	802	802	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	807	807	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CONFLICT	138	138	A -> G (IN REF. 2).	
FT	CONFLICT	614	614	R -> H (IN REF. 3).	
FT	CONFLICT	692	692	P -> L (IN REF. 2).	
SQ	SEQUENCE	930 AA;	101715 MW;	B64281502F28193B	CRC64;

Query Match 43.2%; Score 1745; DB 1; Length 930;
 Best Local Similarity 48.7%; Pred. No. 2.le-119;
 Matches 335; Conservative 98; Mismatches 219; Indels 36; Gaps 10;

QY	2	VSPYRVTLVVADESVMKFGADLEHYLLTLATAARLYRHSPTLNINIVVKKVLLR	61
DB	263	ISRAQVELLVADASMARLYGRGLQHYLLTLASTANRLYSHASTENHRLAVKVVVLG	322
QY	62	DRDSGPKVTGNAALTLNFCNAQKLNKVKDPHEWYTAILETTRDLCGATTCDTGLMA	121
DB	323	DKDKSLEYSKNAATLKNFKWQHONQDDEHYDAAILFTREDLCGHHSCDTLQMA	382
QY	122	DVGTWCDPKRSCSVTEDDGLPSAFTTAHELGHVFNPHDNVVKCEVFGKLRANHMSPT	181
DB	383	DVGTCTSPSCAVIEDDGLHAAFTVAHEIGHLGLSHDDSKFCEETFGSTEDKRLMSI	442
QY	182	LTIQDRANPWSACSAITITFDLDSHGDCLLDQPSKIPSLPDLPGASYTLTQCELAFG	241
DB	443	LTSIDASKPWSKTSATITFEFDLDDGHNCLDLPRKQILGPPELPGQYDATQCNLTFG	502
QY	242	VSGKPCPYMYCTKLCWCTKAKGQWVOTRHPFWADTSCGEGKLCGLKAGACVER---	296
DB	503	PYTSVCGMDCVCAWLCAVVRQGMVCLTKLPVAVGTPCGKGRICLOGKCVDTKKYY	562
QY	297	NLNKHRVDGSKWADPYGPCSRCTCGGVQVLRARQCTNPTPANGKYCEGVRYKRSCLN	356
DB	563	STSSH---GNWGSWGSWGCSSCGGVQVAYRHCNPNAPRNRGYCTGKRAIYRSCSLM	619
QY	357	PCSSASGKSFRECEAFNGYNHSTNRLTLAVANVPKYSVSPKCKLICRANGTVGF	416
DB	620	PCP--PNKGSFRHEQCEAKNGYQSDAKGVKTFVWVPKYAGVLPADVCKLTCKRAGTGY	677
QY	417	YVLAPKVDGTLTSPDSTSVQVQKCIKAGCDNGLSKKRFKDCGCGGDNKSKCKVTGL	476
DB	678	VVFPKVTGTECPYNSVYGVKCVRTGCDIIGSLQYDKCGVCGGDNSSCTKIYGT	737
QY	477	FTKPMHGYNFVWAIPAGASSIDIRQYKGLIGDDNYLALKNSQKYLNLGHFVYSAVER	536
DB	738	FNKSKGYTDVVRIPGEGATHIKVROPKAKDQTRFTAYLALKKKNGEYLINGYIMISTSE	797
QY	537	DLVVGSLRLRYSGTGAVESL-----QASRPILPLTVBVLVSGKMTPPRVYSFVLPE	591
DB	798	IIDINGTVMNSYSGWSHRDDFLHGMGYSAKEL---LIVQILATDTPKPLDVRYSEFVPK	854
QY	592	PREDKSHPKDPRGSPVLHNSVLSL-SNOVEQPDPRPPARVWAGVSGPCASCGSLQKR	650
DB	855	-----STPK-----VNSVTSHGSKNKGSHTSQP--QWVTGPMWACSRCTCDTGWHTR	898
QY	651	AVDCRSAGQRTVPACDAHRPVEVTOAC	678
DB	899	TVQCQ-DGNRKLAKGCPLSQRPSPAFKQC	925

RESULT 9
 AT54_RAT
 ID AT54_RAT STANDARD; PRT; 630 AA.
 AC Q9ESP7; Q9ESP8; Q9ESP6;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
 DE (fragment).
 GN ADAMTS4.

Rattus norvegicus (Rat).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A.
STRAIN=Wistar; TISSUE=Brain;
MEDLINE=20415831; PubMed=10961658;
Satch K., Suzuki N., Yokota H.;
RA "ADAMTS-4 (a disintegrin and metalloproteinase with thrombospondin
RT motifs) is transcriptionally induced in beta-amyloid treated rat
RL astrocytes";
CC Neurosci. Lett. 289:177-180(2000).
CC -! FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGREGAN IN ARTHROITIC DISEASES.
CC -! CATALYTIC ACTIVITY: cleaves aggrecan at the 392-Glu-|-Ala-393
CC site.
CC -! COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -! SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -! TISSUE SPECIFICITY: ONLY EXPRESSED IN BRAIN.
CC -! DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -! PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -! SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -! SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; AB042272; BABI6474.1; -;
DR EMBL; AB042271; BABI6473.1; -;
DR EMBL; AB042273; BABI6475.1; -;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Reprlysins.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_MTPeptdse.
DR Pfam; PF00090; Tsp_1; 2.
DR Pfam; PF01421; Reprlysins; 1.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSPI; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
KW hydrolase; Metalloprotease; zinc; Glycoprotein; Zymogen;
Extracellular matrix.
FT NON_TER 1
FT PROPEP <1 5
FT CHAIN 6
FT METAL 154 154
FT ACT_SITE 155 155
FT METAL 158 158
FT METAL 164 164
FT DOMAIN 233 303
FT DOMAIN 316 367
FT DOMAIN 368 478
FT DOMAIN 479 630
FT DOMAIN 40 45
FT CARBOHYD 96 96
FT CARBOHYD 474 474
FT SEQUENCE 630 AA; 68384 MW; 63A28753167C7EF CRC64;

Query Match 42.8%; Score 1730.5; DB 1; Length 630;
Best local Similarity 52.4%; Presd. No. 1.5e-118;
Matches 322; Conservative 86; Mismatches 183; Indels 21; Gaps 9;

QY	122	DVGTCDP	PRSCSVIEDDGLPSAFTTAHELGHVENPHDNVNVKCEVFCFKLRANHHM	SPT	181
QY	122	DVGTCDP	PRSCSVIEDDGLPSAFTTAHELGHVENPHDNVNVKCEVFCFKLRANHHM	SPT	181
Db	383	DVGTICSP	SPERSCAVIEDDGLHFAHTVAHEIHLGLSHDDSKFCENFGTTEDKRLMSI	442	
QY	182	LIGIDRAN	PWSACSAAIITDFLDSHGDCILLOPSKIPISLPEDLPASVYTL	SOQCELAGF	241
Db	443	LTSIDASK	PNKSKTSATITEFLDDGHGNCILLDLPRKQILGPPELPQGYTADQCNLT	FG	502
QY	242	VGSKPCPY	MOYCYTKLMCTCKAKGQCMVQTRHPFWADGTCGGEGKCLKGACV	ER-----H	296
Db	503	PEYSVCP	GMDDVCAFLCAVVRQGMVCLTKPLPAVEGTCGGRVCLQCKVDKTKKKY	Y	562
QY	297	NLNKHRV	DGSMAXKWDYPGCSRTCCGGVOLLARQCTNPTPANGKKYCEGVVRKYRSCNLE	356	
Db	563	STSSH---	GNWGSWGPWGQCSRCGGGVQFAYHCHNPA	RNSGRYCTGKRAIYRSCSVT	619
QY	357	PCPSSAS	GKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPROCKKLICRANGTGYF	416	
Db	620	PCP--PNG	KSFREHQCEAKNGYQSDAKGVKTFEVMYPKYAGVLPAADVCKLTCRAKGTGY	677	
QY	417	YVLAPKV	DGTLTCLSPDSTSCVGVGCKIKAGCCDGNLGSKKREDKCGVCGDGNKSKKVTGL	476	
Db	678	WTFSPKVT	DTECPYNSVVCVRGRCVTRTCDGGLGSKQYDKCGVCGDGNSSCTKIIGT	737	
QY	477	FTKPMHC	YNFWAIPAGASSIDIRQRYKGLIGLDNYPALNKSQGYLLNGHFVWSAVER	536	
Db	738	FNKSKGY	TDVRIPEGATHIKVRPKAKDQTRFPAYLALKKKTGVEILINGKYMISTSET	797	
QY	537	DLVVKGS	LLRYSGTGTAVESL-----QASRPILPLETLVEVLSVGKMTPPRVRYSFYL	PK	591
Db	798	IIDINGT	VMNYSGWSHRDDFLHGNGYSATKEI---LIVQILATDPTKALGVRYSFVFPKK	854	
QY	592	PREDKS---	SHPKDPGSPVYLNHSLVLSNOVEQDPRPPARWAGVSGWSPCSACSGSLQ	648	
Db	855	TTQKVN	SVIHSNGKVG---HSTQL-----QWVTGFWLACSRCTCDTGW	896	
QY	649	KRAYDCRS	AGQRTVTPACDAAHREVFETQAC	678	
Db	897	TRTVQCC	-DGNRKLKAGCLLSQRPSAFKQ	925	
RESULT 11					
AT12_HUMAN					
AD	AC	AT12_HUMAN	STANDARD;	PRT;	1593 AA.
IC	AD	P58397;			
DT	DT	15-JUN-2002 (Rel. 41, Created)			
DT	DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	DE	ADAM-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12).			
DE	GN	ADAMTS12.			
OS	OS	Human sapiens (Human).			
OC	OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
ON	ON	NCBI_taxid=9606;			
RX	RX	[1]			
RC	RC	SEQUENCE FROM N.A.			
RC	RC	TISSUE=Fetal lung;			
RX	RX	MEDLINE=21264577; PubMed=11279086;			
RA	RA	Cal S., Argueelles J.M., Fernandez P.L., Lopez-Otin C.;			
RT	RT	"Identification, characterization, and intracellular processing of ADAM-TS12, a novel human disintegrin, with a complex structural organization involving multiple thrombospondin-1 repeats."			
RL	RL	J. Biol. Chem. 276:17932-17940(2001).			
CC	CC	-!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).			
CC	CC	-!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).			
CC	CC	-!- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely expressed in gastric carcinomas and in cancer cells of diverse origin.			
CC	CC	-!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY			

CC CC SIMILARITY).

CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.

CC -1- PTM: IS SUBJECT TO AN INTRACELLULAR MATURATION PROCESS LEADING

CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE

CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND

CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1

CC DOMAINS.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -1- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: AJ250725; CAC20419.1; -

CC Genew: HGNC:14605; ADAMTS12.

CC MIM: 606184; -

CC InterPro: IPR001762; Disintegrin.

CC InterPro: IPR002870; Pep_M12B_propep.

CC InterPro: IPR001590; Repolysin.

CC InterPro: IPR000884; TSP1.

CC InterPro: IPR000130; Zn_M12peptdse.

CC Pfam: PF00090; tsp-1; 6.

CC Pfam: PF01421; Repolysin; 1.

CC Pfam: PF01562; Pep_M12B_propep; 1.

CC SMART: SM00209; TSP1; 8.

CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.

CC PROSITE: PS00425; ADAM_MEPRO; 1.

CC PROSITE: PS00426; TSP1; 2.

CC PROSITE: PS00427; TSP1; 2.

CC PROSITE: PS00428; TSP1; 2.

CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

CC Repeat; Extracellular matrix.

CC

CC SIGNAL 1 25

CC FT PROPEP 26 240

CC FT CHAIN 241 1593

CC FT DOMAIN 465 544

CC FT DOMAIN 545 596

CC FT DOMAIN 597 700

CC FT DOMAIN 701 826

CC FT DOMAIN 827 881

CC FT DOMAIN 886 943

CC FT DOMAIN 947 995

CC FT DOMAIN 996 1315

CC FT DOMAIN 1316 1364

CC FT DOMAIN 1367 1423

CC FT DOMAIN 1426 1471

CC FT DOMAIN 1426 1471

CC FT DOMAIN 302 305

CC FT SITE 208 208

CC FT METAL 392 392

CC FT ACT_SITE 393 393

CC FT METAL 396 396

CC FT METAL 402 402

CC FT CARBOHYD 105 105

CC FT CARBOHYD 125 125

CC FT CARBOHYD 215 215

CC FT CARBOHYD 485 485

CC FT CARBOHYD 685 685

CC FT CARBOHYD 790 790

CC FT CARBOHYD 951 951

CC FT CARBOHYD 1104 1104

CC FT CARBOHYD 1275 1275

CC FT CARBOHYD 1300 1300

CC FT CARBOHYD 1320 1320

CC FT CARBOHYD 1371 1371

CC FT CARBOHYD 1378 1378

CC FT CARBOHYD 1503 1503

CC SEQUENCE 1593 AA; 177545 MW; 07F9F4B63BD83A3 CRC64;

Query Match 29.3%; Score 1184.5; DB 1; Length 1593;

Best Local Similarity 34.4%; Pred. No. 2.6e-78;

Matches 279; Conservative 122; Mismatches 280; Indels 129; Gaps 29;

QY 2 VSIPRYVELTVADSMVKFHGAD-LEHYLLTLLATAARLYRHPHSILNFINVVKVLL 60

DB 242 ISKERWVETLVVADTKMIEYHGSENVESYILTIMNWTGLFHNPSIGNAHIVVVRILL 301

QY 61 RDRDSGPKVTGNAALTLRNFCWQKLNKVSQDKHPEYWDTAILFTRODLCA--TTCDTL 118

DB 302 EEEQGLKIVHAEKTLSSFCWQKQKSIKPKSOLNPHVDVAVLLTRKDCAGENRPEC 361

QY 119 GNADVTGMDCKPRSCSVIEDDGLPSAFTAHGLHVFNMHPDNVY-VCEEVFGKLANHM 177

DB 362 GLSHSGMCOPIHRSNINEDSLPLAFTIAHGLHSGFIQHGKENDCEPVG---RHPIY 418

QY 178 MSPTLTIDRANPWASACSAIITDFLDGSHGDCILDDQPSKP-ISPEDLPGASYTL 236

DB 419 MSRLQYDPTPLTWSKSEYITRELDRGWGFCLDDIPKGLKSKVTPAGVIYDVHQC 478

QY 237 ELAFGVSRPCPYMQ-YCTKLWCTGKAGQWVCQTRHPFWADGTSCGEGKCLKGACVER 295

DB 479 QLOYGNATFCQEVENVCTLWCYKG---FCRSKLDAAADGTQCGEKKWCMAGKCI-- 532

QY 296 HNLNK--HRVDGSAKWDYPCSRCTCGGVOLARRQCTNPTPANGKYCGEVVKYRSC 353

DB 533 -TVGKAPESIPGGWGRWSPWCHSCRTCCGAGVQSAERLCNNPEPKGGYCYTGERKRYELC 591

QY 354 NLEPCPSSASGKSFREEQCAFNGYNHSTNRLTLAVAVPKYSGVSPRDKCLICRANGT 413

DB 592 NVHPCRSEA--PTFROMQCEFTVPIYK-NEL---YHWPPIF--NPAPHCELYCRPIDG 642

QY 414 GYFVVLAPKVVDTLC--SPDSTSVCVQKCIKAGCDNLGSKKREDKCGVCGDKNCK 471

DB 643 QFSEKMLDAVIDGTPTCFEGGSRNRCINGICAMVCGDEIDSNATEDRCVCLGDSGSCQ 702

QY 472 KVTGLF-TKPMHGFNVFAIPAGASSIDIRQRYKGLIGDDNLYALKNSQ-GKYLNGHF 529

DB 703 TVRKMEKQEGSGYVDIGLIPKARDIRYME-----IEGAGNFLAIRSEDEPKYYLNGF 757

QY 530 VVSAVERDLVVKGLSLRYSGTGTAVESLOASRPILPTVEVLVSVGKMTTPRVYSFLP 589

DB 758 IIO-WNGNYKLACTVFOYDRKGD-LEKLMATGPTNESVMIIQLLF--QVTNPGIRYETIQ 813

QY 590 KEPREDKSSHPKDPGRGFSVLHNSVLSLSNQVEQ----- 622

DB 814 KD-----GLDNDVEQMYFWQYGHWTGECSTVCTGCTGIRQTAHC 850

QY 623 -----PDRP-----PARVWAGSWGPCASCG-SGLQKRAVD 655

DB 851 IKKGRGMVKATFCDPETOPNGRQKKCHERACPPRWAGEWEACATCGPHGKRTVLICI 910

QY 656 GS--AGORTVPACDAAH--RPVETOACGEP--CPT-WELSAWSPCKSGRGFORRLSKC 708

DB 911 QTVWSDEQALPTDCCOHLKPKTLTLLSCNRDILCPSDWTGVNWSSECVSGGVRIRSVTC 970

QY 709 VGHGRLLRDQCNLHRKPOELDFCVLRPC 738

DB 971 AKNHD-----EPCDVTRKPNRSLCGLQOC 995

RESULT 12

AT10_HUMAN STANDARD; PRT; 1077 AA.

ID AT10_HUMAN

AC Q9H324;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase

GN with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).

DE ADAMTS10.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Apte S.S.;
RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
thrombospondin type 1 repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: AF163762; AAC35563.1; .
DR MEROPS: M12.235; .
DR Genew: HGNC:13201; ADAMTS10.
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR000130; Zn_M12Bptdse.
DR Pfam: PF00090; tspl_1; 5.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR SMART: SM00209; TSP1; 5.
DR PROSITE: PS02015; ADAM_MEROPS; 1.
DR PROSITE: PS00092; TSP1; 2.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT NON_TER 1
FT PROPEP <1 207
FT CHAIN 208 1077
FT METAL 366 366
FT ACT_SITE 367 367
FT METAL 370 370
FT DOMAIN 376 376
FT DOMAIN 434 520
FT DOMAIN 578 679
FT DOMAIN 680 802
FT DOMAIN 821 577
FT DOMAIN 799 860
FT DOMAIN 862 918
FT DOMAIN 922 976
FT DOMAIN 981 1031
FT CARBOHYD 64 64
FT CARBOHYD 196 196
FT CARBOHYD 297 297
FT CARBOHYD 714 714
FT CARBOHYD 769 769
FT CARBOHYD 866 866
SQ SEQUENCE 1077 AA; 118072 MW; 3914DE18DCBBF587 CRC64;
Query Match 28.6%; Score 1155.5; DB 1; Length 1077;
Best Local Similarity 34.0%; Pred. No. 2.le-76;
Matches 291; Conservative 101; Mismatches 304; Indels 159; Gaps 27;
QY 2 VSIPRYVETLVVADSMVKFHG-ADLEHYLLTLLATAARLYRHPHSILNPINIVVKKVLL 60

Db 209 VSRERYVETLVVADSMVKFHG-ADLEHYLLTLLATAARLYRHPHSILNPINIVVKKVLL 268
QY 61 RDRDSGPKVTGNAALTNRNFCAMQKLL-----NKVSKHPEYWDYTAITLFRQLDLC--G 111
Db 269 TEDQPTLEITHHAGSLDSFCWKQKSYVNHSHGNAIPENGVAHDTAVLTRYDICIYK 328
QY 112 ATTCDTLGMADYGTMCDDPKRSCSIEDDGLPSAFTTAHELGHVFNHNDNVKVCVEFG- 170
Db 329 NKPCGTLGLAPVGGMERCERSCSVNEDIGLPOAFTIAHEIGHTFGMNHGVGNSCGARG 388
QY 171 ---KLRAHMMSPFTLIQIDRANP--WSACSAALITDFDSGHGDCLLDQPSK-PISLPED 224
Db 389 DPAKLMAAHITMKT-----NPFVWSSCNRYIITFSDSGLGCLNRRPPQDFVYFV 441
QY 225 LPGAATYLSQCELAFGVGSKPCPYMAYQYTKLWCTGKARGQMVQCOTRHPFADWGTSCGEG 284
Db 442 AFGQAVDADQCFQHGKVSROCKYGEVCSLWCLSKSNR---CITNSIPAAEGTLCQTH 498
QY 285 KL-----CLKGACVERHNLNKHVRDGSWAKWDYGPSCRTCGGGVQLARRQCTNPANGG 340
Db 499 TIDKGWYKRVCPV-FGSRPEGVGAWGPWTWGDGCSRCTCGGVSSSSRHCDSPRPTGG 557
QY 341 KYCEGVYRVYRSCNLEPCPSSASGSRFEEQCEAFNG-----YNHSTNRLTLAVAWVP 393
Db 558 KYCIGERRRHSNTDCCPPGS--QDFREVQCSEFDSIFRKGFKYKWKTYR----- 606
QY 394 KYSGVSPRDCKLICRANGTYFYVLAPKVPDGLTCLSPDSTSVCVQGGKCIKACGDNGLGS 453
Db 607 -GGGVK---ACSLTSLAEGFNFYTERAAAVDGTPCRPTDVCYSGECKHVGCGRVLGS 662
QY 454 KRFDFKCGVCGGDNKCKKVTGLFT--KPMHGYNFVVAIPAGASSIDIRQGYKGLIGDD 511
Db 663 DLREDKCRVCGGDSACETIEGVFSPASGAGYEDVWIPKGSVHFIQD-----LNLSL 717
QY 512 NYLALKNSQGYLLNGHFVYSAVERDLVYKGSLLRYSGTGTAVESLQASRPILPJTVEV 571
Db 718 SHLALKDQESLLEGLPGTPQPHR-LPLAGTTFQLRQGPDOVQSLEALGPINASLIVMV 776
QY 572 LSVGKMTPPRVRYSYFLP----- 589
Db 777 LARTEL--PALRYRFNAPTARDLSPPYSHYAPWTKSAQACGSGVQAVECRNQLDSSA 834
QY 590 -----KEPREDK--SSHEKDP----- 619
Db 835 VAPHYCSAHSKLPKQRQACNTEPCPDVWVGNWSLCSRCDAGVRSRVSVCORRVSAAEE 894
QY 620 VEQ-----PDRPPA-----RWAGSWGSPCSACGSLQKRAVDCRGSAGRTV 663
Db 895 KALDDSSACQPPRPVLEACHGTCPPPEWAALDWSECTPCGPGLRHVRVVLCKSADHRATL 954
QY 664 PA--CDAAHRPVTQAG-EPQP--TWELSAWSPCSKSGRGRFRRSLKCVGHGGRLLAR 718
Db 955 PPAHCSFAAKPPATMRCNLRRCPPARVWAGWEGCSAQCGVGRQVRCTSTGQ--AS 1012
QY 719 DQCNLRHKKPQELDFC 733
Db 1013 HECTEALRPPTTQOC 1027
RESULT 13
AT57_HUMAN STANDARD; PRT; 997 AA.
ID Q0UKF4; 2001 (Rel. 40, Created)
AC Q0UKF4; 2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7).
GN ADAMTS7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99395124; PubMed=10464288;
RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS5, ADAM-TS6, and ADAM-TST, Novel Members of a New Family of
RT Zinc Metalloproteases.";
RL J. Biol. Chem. 274:25555-25563(1999).
CC -|- FACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -|- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG,
CC LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS.
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -|- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licensee@isb-sib.ch).
CC -----
EMBL; AF140675; AAC56358.1; -.
DR HSPP; P15167; IATL.
DR MEROPS; M12.231; -.
DR Genew; HGNC:223; ADAMTS7.
DR MIM; 605009; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn.MTpeptidse.
DR Pfam; PF00090; tsp.1; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPI; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSPI; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DK Hydrolase; Metalloprotease; zinc.
KW Repeat; Extracellular matrix.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 232 BY SIMILARITY.
FT CHAIN 233 997 ADAMTS-7.
FT SITE 204 204 CYSTEINE SWITCH (POTENTIAL).
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 389 389 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 462 537 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 538 594 DISINTEGRIN-LIKE.
FT DOMAIN 595 697 TSP TYPE-1 1.
FT DOMAIN 698 914 CYS-RICH.
FT DOMAIN 915 990 SPACER.
FT CARBOHYD 94 94 TSP TYPE-1 2.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED02FC104 CRC64;

Query Match 27.1%; Score 1094; DB 1; Length 997;
Best Local Similarity 37.5%; Pred.No. 5.8e-72;
Matches 280; Conservative 90; Mismatches 301; Indels 76; Gaps 27;

```

RA Wertelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,
RA Byers P.H., Lapiere C.M., Prockop D.J., Nusgens B.V.;
RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis
are caused by mutations in the procollagen I N-proteinase gene.";
RL Am. J. Hum. Genet. 65:308-317(1999).
CC -1- FUNCTION: Cleaves the propeptides of type I and II collagen prior
to fibril assembly. Does not act on type III collagen. May also
play a role in development that is independent of its role in
collagen biosynthesis.
CC -1- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains
at Ala-1-Gln.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
COLLAGEN TYPE XIV (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (by similarity).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI;
ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-
PROCOLLAGEN PEPTIDASE ACTIVITY.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON
AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
SIMILARITY).
CC -1- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos
syndrome type VIIC (EDS-VIIC), a recessively inherited connective-
tissue disorder characterized clinically by severe skin fragility
and joint hypermobility and biochemically by the presence in skin
of procollagen incompletely processed at the N-terminus.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AJ003125; CAA05880.1; -
DR MEROPS; M12.301; -
DR Genew; HGNC:218; ADAMTS2.
DR MIM; 604539; -
DR MIM; 225410; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTPeptidse.
DR Pfam; PF00090; TSP_1; 4.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 4.
DR PROSITE; PS02015; ADAM_MEROPS; 1.
DR PROSITE; PS00092; TSP1; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;
KW Alternative splicing; Ehlers-Danlos syndrome.
FT SIGNAL 1 29
FT PROPEP 30 253
FT CHAIN 254 1211
FT METAL 408 408
FT ACT_SITE 409 409
FT METAL 412 412
FT METAL 418 418
FT DOMAIN 480 560
BY SIMILARITY.
ADAMTS-2.
ZINC (CATALYTIC) (BY SIMILARITY).
BY SIMILARITY.
ZINC (CATALYTIC) (BY SIMILARITY).
ZINC (CATALYTIC) (BY SIMILARITY).
DISINTEGRIN-LIKE.

FT	DOMAIN	561	617		TSP TYPE-1 1.
FT	DOMAIN	618	722		CYS-RICH.
FT	SITE	691	693		CELL ATTACHMENT SITE (POTENTIAL).
FT	DOMAIN	723	851		SPACER.
FT	DOMAIN	852	911		TSP TYPE-1 2.
FT	DOMAIN	912	974		TSP TYPE-1 3.
FT	DOMAIN	975	1030		TSP TYPE-1 4.
FT	DOMAIN	40	43		POLY-ALA.
FT	DOMAIN	185	188		POLY-GLU.
FT	CARBOHYD	112	112		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	251	251		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	949	949		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	993	993		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1031	1031		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1098	1098		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1145	1145		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1150	1150		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	544	566		HCFGHGIWLTPTDILKRDGWSGA -> FRPGAVAHACYP
FT	VARSPLIC	567	1211		TLGGQGRWIA (IN ISOFORM SPNPI).
FT	VARSPLIC	567	1211		MISSING (IN ISOFORM SPNPI).
FT	SEQUENCE	1211	AA; 134722	MM; BECEE25C323CAD2D	CRC64;
QY	Query Match		26.6%	Score 1075; DB 1; Length 1211;	
Db	Best Local Similarity		34.2%	Pred. No. 1.7e-70;	
QY	Matches 276; Conservative 110; Mismatches 300; Indels 120; Gaps 35;				
QY	8	VETLVVADESMVKFGAD-LEHYLTLTATARLYRHPHSILNPINIVVVKVLLLRDRDSG	66		
Db	268	IEVLLGVDDSVVQFHGKEHVQKYLTLNMVINEIYHDESLGAHINVLVILLISYKSKM	327		
QY	67	PKV-TGNALTRNCAMQKLNKVSDEKHPYWDATILFTQDLQCGATTCTGLMADYGT	125		
Db	328	STIEIGNPSQSLNCRWAYLOOKPDTGDEYHDAIFLTRQDF-GPSGMO--GVAPVTG	384		
QY	126	MCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHNV-KVCEEVEGKLRANHMSPTLIQ	184		
Db	385	MCHPVRSCTLNHEDGFSSAFVVAHETGHVLMGHEHQGNRCGD---EVLRLSINAPLVQA	441		
QY	185	IDRANPWSACSAITDLDLSDGHCILLDP---SKPTSLPDLPGASYTLQOCELAFG	241		
Db	442	AFRHFHWSRCSQOELSLYLS--DCLDDDPFAHDWP-ALPO-LPLGLHYSMNEQCRFDFG	497		
QY	242	VGSKPCPYMQY---CTKLWCTGKAKQVMVQTRHFFPWADGTCGEGKLCGLKACV----	293		
Db	498	LGVMCTAFRTDPCKQLWCS-HPDNPYCKTKKGPPLDGTWCAPGKCFKHCILWLPD	556		
QY	294	--ERHNLNKHVRVDGSAKWWDYPGCSRTCCGGVQLARQCTNPTPANGKCYEGVRYKR	351		
Db	557	ILKR-----DGSWGAWSPEFGSCSRTCGTGKFTQCDNPHFPANGRTCSGLAYDFQ	608		
QY	352	SNLEPCPSASGKSPREECEAFNGY-----NHSTNRLTLAVAVWPKYGVSPROCK	404		
Db	609	LCSRQDCPDSLA--DFRESQCRQWLDLYFEHGDAQHH-----WLP-HEHRDAKER	655		
QY	405	KLICRANGTGYFYVLAPKVVDGTLCS-PDSTSVGVQKCIKAGCDNGLGSKRFRKCGVC	463		
Db	656	HLVCESRETGEVVMKRMVHDGTRCSYKDAFSLCVRGDCRKYGCDGVIGSSQEDKCGVC	715		
QY	464	GGDNKSKKVTGLFTK--PMHGYNFVVAIPAGASSIDIRQGYKGLIGDNDYALKNSQ-	520		
Db	716	GGDNSHCKVYKGTFTSRPKKHGYIKMFEIPAGARHLIIQE-----VDATSHHLAVKNLET	770		
QY	521	GKYLNGHFVVSVERDLVYKGLSLRAYSGTGTAVESLQASRPLTLEPLTVLSVGMKTPP	580		
Db	771	GKFIENEVDNDASSKTFTAMGVENEYRDE-DGRETLQTMGLPHGTITVLVTPVGD-TRV	828		
QY	581	RVRYSYLPKEP-----RED-----KSSHPKDPGR-----PS	607		
Db	829	SLIYKMIHEDSLNVDNNVLEEDSVVYEWALKKSPCKSGGGGSOFTKYGCRRLDHK	888		
QY	608	VLHN---SVLSLSNVEQ---PDRPRPARWAGSWGPGCSAGC-SGLQKRAVDCRSAGQ	660		
Db	889	MVHRGFCALSKPKATIRACNPOECSPQVVMVTGEWEPSCOTCGRTGQVRSVRCIOPLHD	948		

Fri May 2 10:41:54 2003

```
QY 473 VTGLTKPMH--GYNFVAIPAGASSIDIRQGYKGLIGDD----NYLAKN-SQKVI 525
Db 715 VKGTFTTRPKUGYLNKFDIPPGARHV-----LIQDEASPHILAIAKQATGHYIL 765
QY 526 NGH-----FVVSVERDLVVKSLRLRYSGTGTAVESLQASRPLEPLTVEVLVSGKM 577
Db 766 NGKGEAKSRTFDLGVENDYNIEDD-----IESLHTDGPLHDPVILIIQEND 815
QY 578 TPRVRYSFYLPKEPREDKSSHPKDPGRGVSVLHNSV-----LSLSNQVEOP--- 623
Db 816 TRSSLTYKYII---HEDSV-----PTINSNNVIQEEELDTFEWALKSWQVSKPCGG 863
QY 624 -----DDRP-PAR-----WVAGSWGPCASCG- 644
Db 864 GFQYKYGCRKSDNKMVHRSCFCEANKKXPPIRMCMNIQECIHPWVABEWEHCTKTCGS 923
QY 645 SGLQKRAVDCRGAGQRTVPACDAHRPVETQAC-GE-----PCPT-WELSAWS 691
Db 924 SGYQLRTVRC-----LQPLDGTNRSHVSKYCMGDRPESRRPCNRVPCPAQWKTPWS 976
QY 692 PCSKSCGRGFORSLKCVHGGRLARDQCNLHRKPQELDFCVLRPC 738
Db 977 ECSVTCGEGTEVRQVLC-----RAGDHCD-GEKPESVRACQLPPC 1015
```

Search completed: April 29, 2003, 17:14:25
Job time : 26.0892 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2003, 19:21:51 ; Search time 7406 Seconds
(without alignments)
11211.221 Million cell updates/sec

Title: US-10-009-332-2

Perfect score: 2853

Sequence: 1 atgtctttgtggcatcct.....gcgtctgagcgctgctga 2853

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2848.2	99.8	2853	9	HSA315733	AJ315733 Homo sapi
2	2814.4	98.6	2937	6	AX319860	AX319860 Sequence
3	2812.8	98.6	2930	6	AX342635	AX342635 Sequence
4	957.2	33.6	157963	9	AP002986	AP002986 Homo sapi
5	957.2	33.6	172905	2	AC025130	AC025130 Homo sapi
6	893	31.3	170682	2	AC023429	AC023429 Homo sapi
7	779.8	27.3	182656	2	AC101990	AC101990 Mus muscu
8	778.8	27.3	28000	9	AP003459	AP003459 Homo sapi
9	775.8	27.2	178764	2	AC126507	AC126507 Rattus no
10	775.2	27.2	3927	10	BC009667	BC009667 Mus muscu
11	666	23.3	2670	6	E55282	E55282 Novel metal
12	666	23.3	2670	6	E58655	E58655 Novel metal
13	666	23.3	3711	9	AF060153	AF060153 Homo sapi
14	636	22.3	4180	10	D67076	D67076 Mouse mRNa
15	631.4	22.1	4659	9	AF207664	AF207664 Homo sapi
16	629.8	22.1	3430	9	AF060152	AF060152 Homo sapi
17	629.8	22.1	4014	6	E29406	E29406 Novel integ
18	629.8	22.1	4447	9	AF170084	AF170084 Homo sapi
19	628.8	22.0	2184	6	E29966	E29966 Human ADAMT
20	628.2	22.0	2853	6	E58656	E58656 Novel metal
21	628.2	22.0	4670	9	BC036515	BC036515 Homo sapi
22	626.6	22.0	4309	9	AB037767	AB037767 Homo sapi
23	614.4	21.5	4878	10	AF149118	AF149118 Rattus no
24	608	21.3	2857	10	AF304446	AF304446 Rattus no
25	599	21.0	178764	2	AC126507	AC126507 Rattus no
26	593.6	20.8	3638	10	AF175282	AF175282 Mus muscu
27	584.8	20.5	4193	9	AF148213	AF148213 Homo sapi
28	584.8	20.5	4301	9	AB014588	AB014588 Homo sapi
29	583.2	20.4	4407	6	AX376284	AX376284 Sequence
30	564.2	19.8	3321	10	BC027773	BC027773 Mus muscu
31	547.4	19.2	3002	10	AF140673	AF140673 Mus muscu
32	547.4	19.2	5533	9	AF142099	AF142099 Homo sapi
33	419.2	14.7	5139	9	AB037733	AB037733 Homo sapi
34	419.2	14.7	5808	6	AX319857	AX319857 Sequence
35	417.6	14.6	3674	9	AF261918	AF261918 Homo sapi
36	404	14.2	1251	10	AB042271	AB042271 Rattus no
37	384	13.5	2838	9	HSM802617	AL162080 Homo sapi
38	377.2	13.2	5724	6	AX392458	AX392458 Sequence
39	377.2	13.2	6060	6	AX392460	AX392460 Sequence
40	374	13.1	172905	2	AC025130	AC025130 Homo sapi
41	366.2	12.8	1560	9	AF141293	AF141293 Homo sapi
42	322	11.3	2554	9	AK094057	AK094057 Homo sapi
43	321.2	11.3	1365	10	AB042272	AB042272 Rattus no
44	314.4	11.0	3308	6	AX480930	AX480930 Sequence
45	314.4	11.0	4518	6	AX319858	AX319858 Sequence

ALIGNMENTS

RESULT 1	HSA315733	Homo sapiens mRNA for metalloprotease disintegrin 15 (ADAMTS15 gene)	2853 bp	mRNA	lineae	FRI 01-MAR-2002
LOCUS	AJ315733					
DEFINITION	ADAMTS15					
ACCESSION	AJ315733.1					
VERSION	GI:19171175					
KEYWORDS	ADAMTS15 gene; disintegrin; metalloprotease; thrombospondin.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1					
AUTHORS	Cal.S., Obaya,A.J., Llamazares,M., Garabaya,C., Quesada,V. and Lopez-Ortin,C.					

TITLE		Cloning, expression analysis, and structural characterization of seven novel human ADAMTS, a family of metalloproteinases with disintegrin and thrombospondin-1 domains	
JOURNAL MEDLINE	21856482	Gene 283 (1-2), 49-62 (2002)	
PUBMED	11867212		
REFERENCE	2 (bases 1 to 2853)		
AUTHORS	Cal,S.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-JUN-2001)	Cal S., Biochemistry and Molecular Biology, University of Oviedo, Campus del Cristo, Asturias. 33006, SPAIN	
FEATURES	Location/Qualifiers		
source	1..2853	/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
gene	1..2853		
		/gene="ADAMTS15"	
CDS	1..2853	/gene="ADAMTS15"	
		/codon_start=1	
		/product="metalloprotease disintegrin 15 with thrombospondin domains"	
		/protein_id="CAC86014.1"	
		/db_xref="GI:19171176"	
		/translation="MLLGLTTLTFAFAGTAGGSEPEREVVPIRLDPLDINGVRYWRGPEPSGGDLFLQITAFQEDPFLHLTPADQFLAPAFSTEHLVPLQGLTGGSSDLRRFCYSGDVNAEPDSFAVSLCGGLRCAFGVGAEEVVISPLPNASAPAAORNSQGAHLQRRGVGGPSGDDPTSCGVASGNPAILRALDYPKPRRAGFGESSRRSRGVRVPSIPRYVETLVADSESRMGFHGADLHETLLTLATAARLYRHPISILNPINIVKRVLLRRDSSGPTKCPNSALTIURNECAWGLKFNKXDKHPEYWDTAILFTQDLGCGATCTDGLGNADTLQIDRANPSCASAAIITFDLSDGSHGCVLLDQPSKPSLPEDLPASVYTLSSQOCELAFVGGKPCPMOYCTKLMCTGKAKGMVOLTQRPFPADGTSCGEGKCLKGACVERHLNKNHRVDGSAWMDPVGPCRSRTCGGVOLARQCTNPTPANGKYCEGVRVKYRSCNLEPCPSAAGKSFREEQCAFNGYNHSTNRLFLAYAVPKYSGVSRPKCKLCILRANGTGFEYVLA PKVDGTLCSDDTSVCVCGKCIKAGCDNGLSKRRFPDKCGVCGDNNKSKVYIVLTKPMHGNRYNVAIPAGASSIDIRQGYKGLIGDDNYLAKNSQGYLLNHGFVYSVERDLVLRYSGLTSGTAVESLQASRPLEITVEVLSVGMKTPPRVRYSFYLPKEPREDKSHHPKDPGSPVLHNSVLSLNSQVEQDPRDPARWAWGSGWPCASACGSLGQKRAVDCRGAGQRTQVPACDAHRPVEVTOACGPECPTELWELSAWSPCKSCGRGFGQRRSLKCVGHGGRLLARDQCNLHRKPKQLDFCVLRPC"	
BASE COUNT	522 a	921 c	903 g 507 t
ORIGIN			
Query Match	99.8%	Score 2848.2	DB 9; Length 2853
Best Local Similarity	99.9%	Pred. No. 0;	
Matches 2850;	Conservative	0; Mismatches	3; Indels 0; Gaps 0;
Qy	1	ATGCTTTTGTCTGGCATCTTAACCCCTGGCTTCGCGGGCGAACCCTGAGGCTTTGAG	60
Db	1	ATGCTTCTGTCTGGCATCTTAACCCCTGGCTTCGCGGGCGAACCCTGAGGCTCTGAG	60
Qy	61	CCAGAGCGGGAGTGTCTGCCATCCGACTGCAGCCGCGACATTAACGGCCGCGCTAC	120
Db	61	CCAGAGCGGGAGTGTCTGCCATCCGACTGCAGCCGCGACATTAACGGCCGCGCTAC	120
Qy	121	TACTTGGCGGGTCCGAGGACTCGGGGATCAGGACTCATTTTTCAGATCACAGCATTT	180
Db	121	TACTTGGCGGGTCCGAGGACTCGGGGATCAGGACTCATTTTTCAGATCACAGCATTT	180
Qy	181	CAGGAGGACTTTTACCTPACACCTGCAGCCGGATGCTCAGTTCTTGGCTCCCGCTTCTCC	240
Db	181	CAGGAGGACTTTTACCTPACACCTGCAGCCGGATGCTCAGTTCTTGGCTCCCGCTTCTCC	240
Qy	241	ACTGAGCATCTGGCGGTCCCTCCAGGGGCTCACCGGGGGCTCTTTCAGACCTGCGACGC	300
Db	241	ACTGAGCATCTGGCGGTCCCTCCAGGGGCTCACCGGGGGCTCTTTCAGACCTGCGACGC	300
Qy	301	TGCTTCTATTCTGGGACGTGAACCGGACCGGACTCGTTCGCTGTGAGCCCTGTGC	360
Db	301	TGCTTCTATTCTGGGACGTGAACCGGACCGGACTCGTTCGCTGTGAGCCCTGTGC	360
Qy	361	GGGGGGCTCCCGGAGGCTTTTGGCTACCGAGGCGCGGAGTATGTCTATTAGCCCGCTGCC	420

```
Db 1441 CAGACCCGCACTTCCCTCGGGCGGATGGCACCAGCTGTGGCGAGGCAAGCTCTGCCTC 1500
QY 1501 AAAGGGCCCTCGCTGGAGACACAACTCAACAACACAGSGTGGATGGTTCTTGGGCC 1560
Db 1501 AAAGGGCCCTCGCTGGAGACACAACTCAACAACACAGSGTGGATGGTTCTTGGGCC 1560
QY 1561 AAATGGGATCCCTATGAGCCCTCTGCTGGCGACATGTGGTGGGGCGCTGCAGCTGGCCAG 1620
Db 1561 AAATGGGATCCCTATGAGCCCTCTGCTGGCGACATGTGGTGGGGCGCTGCAGCTGGCCAG 1620
QY 1621 AGGAGTGCACCAACCCCTCTGCAAGGGGGGAAGTACTGCGAGGAGTGGGGTG 1680
Db 1621 AGGAGTGCACCAACCCCTCTGCAAGGGGGGAAGTACTGCGAGGAGTGGGGTG 1680
QY 1681 AAATACCATCTGCAACCTGGAGCCCTGCCAGCTCAAGCTCCCGGAAGAGCTTCGG 1740
Db 1681 AAATACCATCTGCAACCTGGAGCCCTGCCAGCTCAAGCTCCCGGAAGAGCTTCGG 1740
QY 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGCACCACCAACCGGCTCACTCTCGC 1800
Db 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGCACCACCAACCGGCTCACTCTCGC 1800
QY 1801 GTGGCATGGTGGCCCAAGTACTCGGGGTGTCTCCCGGGACAAAGTGCAGCTCATCTGC 1860
Db 1801 GTGGCATGGTGGCCCAAGTACTCGGGGTGTCTCCCGGGACAAAGTGCAGCTCATCTGC 1860
QY 1861 CGAGCCAAATGGCACTGGCTACTTCTATGTGTGGCACCAAGGTGTGGACGCGACGCTG 1920
Db 1861 CGAGCCAAATGGCACTGGCTACTTCTATGTGTGGCACCAAGGTGTGGACGCGACGCTG 1920
QY 1921 TGCTCTCTGACTCCACTCGCTGTGTCTTCAAGGCAAGTGCATCAAGGCTGGCTGTGAT 1980
Db 1921 TGCTCTCTGACTCCACTCGCTGTGTCTTCAAGGCAAGTGCATCAAGGCTGGCTGTGAT 1980
QY 1981 GGGAACTGGCTCCCAAGAGATTCGACAAGTGTGGGTGTGGGGAGGACATAG 2040
Db 1981 GGGAACTGGCTCCCAAGAGATTCGACAAGTGTGGGTGTGGGGAGGACATAG 2040
QY 2041 AGTGTCAAGAGGTGACTGGACTCTTCAACCAAGCCATGCTGGCTACAAATTTCTGGTG 2100
Db 2041 AGTGTCAAGAGGTGACTGGACTCTTCAACCAAGCCATGCTGGCTACAAATTTCTGGTG 2100
QY 2101 GCCATCCCGCAGCGGCTCAAGCATCGACATCCCGCAGCGGTTACAAAGGCTGATC 2160
Db 2101 GCCATCCCGCAGCGGCTCAAGCATCGACATCCCGCAGCGGTTACAAAGGCTGATC 2160
QY 2161 GGGATCACAACTACCTGGCTCTGAAGACAGCAGCAGCAAGTACCTGTCAACGGGAT 2220
Db 2161 GGGATCACAACTACCTGGCTCTGAAGACAGCAGCAGCAAGTACCTGTCAACGGGAT 2220
QY 2221 TTCTGTGTGTGGGGTGGAGCGGACCTGGTGTGAAGGGCAGTCTGTGGGTACAGC 2280
Db 2221 TTCTGTGTGTGGGGTGGAGCGGACCTGGTGTGAAGGGCAGTCTGTGGGTACAGC 2280
QY 2281 GGCACGGGCACAGGGTGGAGAGCCTGAGGCTTCCCGGCCCATCTGGAGCGGCTGACC 2340
Db 2281 GGCACGGGCACAGGGTGGAGAGCCTGAGGCTTCCCGGCCCATCTGGAGCGGCTGACC 2340
QY 2341 GTGGAGTCTCTCTCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTTCTATCTG 2400
Db 2341 GTGGAGTCTCTCTCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTTCTATCTG 2400
QY 2401 CCCAAAGAGCCTGGGAGGACAAGTCTCTCATCCCAAGGACCCCGGGGACCCCTCTGTC 2460
Db 2401 CCCAAAGAGCCTGGGAGGACAAGTCTCTCATCCCAAGGACCCCGGGGACCCCTCTGTC 2460
QY 2461 TTGCACAACAGCTCTCAGCCCTCTCCAAAGAGTGGAGAGCGGACAGAGGCCCT 2520
Db 2461 TTGCACAACAGCTCTCAGCCCTCTCCAAAGAGTGGAGAGCGGACAGAGGCCCT 2520
QY 2521 GCAGGCTGGTGGTGGCAGCTGGGGCGGTGCTCCCGGAGCTGCGGCGAGTGGCCCTGAG 2580
Db 2521 GCAGGCTGGTGGTGGCAGCTGGGGCGGTGCTCCCGGAGCTGCGGCGAGTGGCCCTGAG 2580
```

QY 361 GGGGGGTCGCGAGAGCCTTTGGCTACCGAGCGCGAGTATGTCTCATTTAGCCCGCTGCCCC 420
Db 436 GGGGGGTCGCGAGAGCCTTTGGCTACCGAGCGCGAGTATGTCTCATTTAGCCCGCTGCCCC 495
QY 421 AATGCTAGCGCGCGCGCGCGAGCGACACAGCGCGCGCGACACCTTCTCCAGCGCGCG 480
Db 496 AATGCTAGCGCGCGCGCGCGAGCGACACAGCGCGCGCGACACCTTCTCCAGCGCGCG 555
QY 481 GGTGTTCCGGCGCGCGCTTCCGAGAGCCCGACACCTCTCGCTCGGGGTGCGCTCGGGCTGG 540
Db 556 GGTGTTCCGGCGCGCGCTTCCGAGAGCCCGACACCTCTCGCTCGGGGTGCGCTCGGGCTGG 615
QY 541 AACCCCGCCATCTACGGGCCCTTGAGCCCTTACAAGCGCGCGCGCGCGCTTCCGGGAG 600
Db 616 AACCCCGCCATCTACGGGCCCTTGAGCCCTTACAAGCGCGCGCGCGCGCTTCCGGGAG 675
QY 601 AGTCGTAGCGCGCGAGTCTGCGCGCGCGCAAGCGTTTCTGCTATCCCGGGTACGTTG 660
Db 676 AGTCGTAGCGCGCGAGTCTGCGCGCGCGCAAGCGTTTCTGCTATCCCGGGTACGTTG 735
QY 661 GAGACGCTGTGTGCGCGAGAGTCAATGGTCAAGTTTCCACGCGCGCGAGCTTGAACAT 720
Db 736 GAGACGCTGTGTGCGCGAGAGTCAATGGTCAAGTTTCCACGCGCGCGAGCTTGAACAT 795
QY 721 TATCTGCTGAGCTGTGCGCAAGCGCGCGGCGACTCTACGGCCATCCAGCATCTCTCAAC 780
Db 796 TATCTGCTGAGCTGTGCGCAAGCGCGCGGCGGCGACTCTACGGCCATCCAGCATCTCTCAAC 855
QY 781 CCATCAACATCTGTGTGCAAGTGTCTCTTATAGAGATGCTGACCTCGGCGCGCAAG 840
Db 856 CCATCAACATCTGTGTGCAAGTGTCTCTTATAGAGATGCTGACCTCGGCGCGCAAG 915
QY 841 GTCACCGGCAATCGGCGCTGACGCTGCGCAACTTCTGCTGTGCGAGAGAGCTGAAC 900
Db 916 GTCACCGCAATCGGCGCTGACGCTGCGCAACTTCTGCTGTGCGAGAGAGCTGAAC 975
QY 901 AAGTGAGTGACAGACCCCGAGTACTGGGACACTGCCATCTCTTACCAGGCGAGGAC 960
Db 976 AAGTGAGTGACAGACCCCGAGTACTGGGACACTGCCATCTCTTACCAGGCGAGGAC 1035
QY 961 CTGTGTGGAGCCACCACTGTGACACCTTGGGATGGCTGATGTGGGTACCATGTGTGAC 1020
Db 1036 CTGTGTGGAGCCACCACTGTGACACCTTGGGATGGCTGATGTGGGTACCATGTGTGAC 1095
QY 1021 CCCAAGAGAGCTGCTGTCTATTGAGAGAGATGGGCTTCCATFAGCCTTCAACACTGCC 1080
Db 1096 CCCAAGAGAGCTGCTGTCTATTGAGAGAGATGGGCTTCCATFAGCCTTCAACACTGCC 1155
QY 1081 CAGAGCTGGCGCACGTGTTCAACATGCCCATGACCAATGTGAAGTCTGTGAGGAGTG 1140
Db 1156 CAGAGCTGGCGCACGTGTTCAACATGCCCATGACCAATGTGAAGTCTGTGAGGAGTG 1215
QY 1141 TTTTGGGAAGCTCGAGCAACCACTGATGTCCCGACCTTATCCAGATCGACCGTGGC 1200
Db 1216 TTTTGGGAAGCTCGAGCAACCACTGATGTCCCGACCTTATCCAGATCGACCGTGGC 1275
QY 1201 AACCCCTGTGACGTGCTGCTGATCATCATCACGACTTCTGTGACAGCGCGGACGCT 1260
Db 1276 AACCCCTGTGACGTGCTGCTGATCATCATCACGACTTCTGTGACAGCGCGGACGCT 1335
QY 1261 GACTGCTCTGACCAACCCAGACAGCCATCTCCTGCCGAGGATCTGCCGGCGCC 1320
Db 1336 GACTGCTCTGACCAACCCAGACAGCCATCTCCTGCCGAGGATCTGCCGGCGCC 1395
QY 1321 AGCTACACCTGAGCAGAGTCCGAGCTGGCTTTTGGCGTGGGCTTCCAAAGCCCTGTCT 1380
Db 1396 AGCTACACCTGAGCAGAGTCCGAGCTGGCTTTTGGCGTGGGCTTCCAAAGCCCTGTCT 1455
QY 1381 TACATGAGTACTGACCAAGCTGTGGTACCGGGAAGGCCAAGGACAGATGTGTGC 1440
Db 1456 TACATGAGTACTGACCAAGCTGTGGTACCGGGAAGGCCAAGGACAGATGTGTGC 1515

QY 1441 CAGACCCCGCACTTCCCTGGCGCATGGCACAGCTGTGGCGAGGGCAAGCTCTGCCTC 1500
Db 1516 CAGACCCCGCACTTCCCTGGCGCATGGCACAGCTGTGGCGAGGGCAAGCTCTGCCTC 1575
QY 1501 AAAGGGGCTCGTGGAGAGACAACTCAACAGACACAGGTGGATGTTCTTGGGCC 1560
Db 1576 AAAGGGGCTCGTGGAGAGACAACTCAACAGACACAGGTGGATGTTCTTGGGCC 1635
QY 1561 AAATGGGATCCTATGGCCCTTGTCTGCGCACATGTGTGGGGGCGTGCAGCTGGCCAGG 1620
Db 1636 AAATGGGATCCTATGGCCCTTGTCTGCGCACATGTGTGGGGGCGTGCAGCTGGCCAGG 1695
QY 1621 AGGAGTGCACCAACACCCCTGCGCAACGGGGGCAAGTACTCGGAGGAGTGAAGGTG 1680
Db 1696 AGGAGTGCACCAACACCCCTGCGCAACGGGGGCAAGTACTCGGAGGAGTGAAGGTG 1755
QY 1681 AAATACCATCTGCAACCTGAGCCCTGCGCACAGCTCAGCTCCGGAAGAGTTCGG 1740
Db 1756 AAATACCATCTGCAACCTGAGCCCTGCGCACAGCTCAGCTCCGGAAGAGTTCGG 1815
QY 1741 GAGGAGAGTGTGAGGCTTTCAACGGCTTAAACACAGCACACACCGGCTCACTCTCGCC 1800
Db 1816 GAGGAGAGTGTGAGGCTTTCAACGGCTTAAACACAGCACACACCGGCTCACTCTCGCC 1875
QY 1801 GTGGCATGGTGGCCCAAGTACTCGGGCGTGTCTCCCGGGACAAAGTCAAGCTCATCTGC 1860
Db 1876 GTGGCATGGTGGCCCAAGTACTCGGGCGTGTCTCCCGGGACAAAGTCAAGCTCATCTGC 1935
QY 1861 CGAGCAATGGCACGTGCTTCTATGTCTGCGCACCCAA---GGTGGTGGACGGCAGC 1917
Db 1936 CGAGCAATGGCACGTGCTTCTATGTCTGCGCACCCAAAGTGGTGGAGCGGACG 1995
QY 1918 CTGTGCTCTCTGACTCCACCTCGCTGTGTCCAAGCAAGTGCATCAAGCTTCTGCTG 1977
Db 1996 CTGTGCTCTCTGACTCCACCTCGCTGTGTCCAAGCAAGTGCATCAAGCTTCTGCTG 2055
QY 1978 GATGGAACTTGGCTCCAAAGAGATTCCAAAGTGTGGGGTGTGGGGGAGACAAT 2037
Db 2056 GATGGAACTTGGCTCCAAAGAGATTCCAAAGTGTGGGGTGTGGGGGAGACAAT 2115
QY 2038 AAGAGCTGCAAGAGTGTGACTCTTCAACAGCCCATGSCATGGCTACAATTTCTG 2097
Db 2116 AAGAGCTGCAAGAGTGTGACTCTTCAACAGCCCATGSCATGGCTACAATTTCTG 2175
QY 2098 GTGGCCATCCCGCAGGCGCTCAAGCATCGACATCCGCCAGCGGGTTACAAAGGGCTG 2157
Db 2176 GTGGCCATCCCGCAGGCGCTCAAGCATCGACATCCGCCAGCGGGTTACAAAGGGCTG 2235
QY 2158 ATCGGGGATGACAATCTGCTGCTGAAGAACAGCCAAAGTACCTCTCTCAACGG 2217
Db 2236 ATCGGGGATGACAATCTGCTGCTGAAGAACAGCCAAAGTACCTCTCTCAACGG 2295
QY 2218 CATTTCTGTTGCTCGGGGAGCGGACCTGTTGTTGAAGGCGAGTCTGCTCGGTAC 2277
Db 2296 CATTTCTGTTGCTCGGGGAGCGGACCTGTTGTTGAAGGCGAGTCTCTCGGTAC 2355
QY 2278 AGCGGCAGGGCAGACAGCGGTGAGAGCTGAGGGTTCCCGGGCCCATCTCTGGAGCCGCTG 2337
Db 2356 AGCGGCAGGGCAGACAGCGGTGAGAGCTGAGGGTTCCCGGGCCCATCTCTGGAGCCGCTG 2415
QY 2338 ACCGTGAGGTCTCTCCGTGGGAAGATGACACCGCCCGGGTCCGCTACTCTCTCTAT 2397
Db 2416 ACCGTGAGGTCTCTCCGTGGGAAGATGACACCGCCCGGGTCCGCTACTCTCTCTAT 2475
QY 2398 CTGCCAAAGAGCTCGGGAGGACAAGTCTCTCATCCCAAGGACCCCGCG- ----GGA 2451
Db 2476 CTGCCAAAGAGCTCGGGAGGACAAGTCTCTCATCCCGGCACCCCGGGGAGGAGGA 2535
QY 2452 CCCTCTGTGTGCACACAGCGTCTCAGCTCTTCAACAGGTGGAGAGCCGCGAGAC 2511
Db 2536 CCCTCTGTGTGCACACAGCGTCTCAGCTCTTCAACAGGTGGAGAGCCGCGAGAC 2595
QY 2512 AGGCCCCCTGCACGCTGGGTGGCTGGCAGCTGGGGCGCTGCTCCCGAGCTGGCGAGT 2571

```
Db 2596 AGGCCCCCTGCACGCTGGGTGGCTGGCAGCTGGGGGCGGTCTCCGCGAGCTCGCGCAGT 2655
QY 2572 GGCCTGCAGAAAGGGGGGGTGGACTGCCGGGCTCCGGCGGAGCGCAGCGTCCCTGCC 2631
Db 2656 GGCCTGCAGAAAGGGGGGGTGGACTGCCGGGCTCCGGCGGAGCGCAGCGTCCCTGCC 2715
QY 2632 TGTGATGACGCCATCGCCGCGCTGGAGACACAAAGCCTGGGGGAGCCCTGCCCACTGG 2691
Db 2716 TGTGATGACGCCATCGCCGCGCTGGAGACACAAAGCCTGGGGGAGCCCTGCCCACTGG 2775
QY 2692 GAGCTAGCGCGTGGTACCGTCTCCAGAGCTGCGGGCGGGGATTTTCAGAGCGCTCA 2751
Db 2776 GAGCTAGCGCGTGGTACCGTCTCCAGAGCTGCGGGCGGGGATTTTCAGAGCGCTCA 2835
QY 2752 CTCAAGTGTGTGGCCACGAGCGCGGCTGCTGCGCCGGGACAGTGCACCTTGCACCG 2811
Db 2836 CTCAAGTGTGTGGCCACGAGCGCGGCTGCTGCGCCGGGACAGTGCACCTTGCACCG 2895
QY 2812 AAGCCCGAGAGCTGGACTTCTCGCTCTGAGCCGCTGTGA 2853
Db 2896 AAGCCCGAGAGCTGGACTTCTCGCTCTGAGCCGCTGTGA 2937

RESULT 3
AX342635
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
AUTHORS
Yue H., Elliott, V.S., Gandhi, A.R., Lal, P., Au-Young, J.,
Tribouley, C.M., Deleane, A.M., Baughn, M.R., Nguyen, D.B., Lee, E.A.,
Hafalia, A., Khan, F.A., Walla, N.K., Yao, M.G., Lu, D.A., Patterson, C.,
Tang, Y.T., Walsh, R.T., Azimzai, Y., Ramkumar, J., Xu, Y. and Reddy, R.
JOURNAL
Patent: WO 0198468-A 32 27-DEC-2001;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
Source
1..2930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 7473089CBI"
BASE COUNT
529 a 958 c 931 g 512 t
ORIGIN
Query Match 98.6%; Score 2812.8; DB 6; Length 2930;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2843; Conservative 0; Mismatches 7; Indels 6; Gaps 2;
QY 1 ATGCTTTTCTGGGATCTTAACCTTGGCTTTCGGCGGCGAACGCTGGAGCTTTGAG 60
Db 75 ATGCTTCTGCTGGGATCTTAACCTTGGCTTTCGGCGGCGAACGCTGGAGCTTTGAG 134
QY 61 CCAGAGCGGGAGGTAGTCTGCTCCCTCCAGCTGGACCGGACATTAAGCGCGCGCTAC 120
Db 135 CCAGAGCGGGAGGTAGTCTGCTCCCTCCAGCTGGACCGGACATTAAGCGCGCGCTAC 194
QY 121 TACTGGCGGGTCCCGAGAGCTCGGGGATCAGGACTCATTTTTCAGATCACAGATTT 180
Db 195 TACTGGCGGGTCCCGAGAGCTCGGGGATCAGGACTCATTTTTCAGATCACAGATTT 254
QY 181 CAGGAGGACTTTTACTACACTGACGCGGATGCTCAGTTCTGCTCCGCGCTTCC 240
Db 255 CAGGAGGACTTTTACTACACTGACGCGGATGCTCAGTTCTGCTCCGCGCTTCC 314
QY 241 ACTGAGCATCTGGGCTCCCGCTCCAGGGGCTCACCGGGGCTCTTTCAGACTCGGACGC 300
Db 315 ACTGAGCATCTGGGCTCCCGCTCCAGGGGCTCACCGGGGCTCTTTCAGACTCGGACGC 374
```

```
QY 301 TGCCTTCTATTCTGGGGACGCTGAACCGCGAGCCGAGCTCGTTCCGCTGCTGTGAGCTGTGC 360
Db 375 TGCCTTCTATTCTGGGGACGCTGAACCGCGAGCTCGTTCCGCTGCTGTGAGCTGTGC 434
QY 361 GGGGGCTCCGCGGAGCCCTTTGGCTACCGAGGCGCGGAGTATGTCTATTAGCCGCTGCC 420
Db 435 GGGGGCTCCGCGGAGCCCTTTGGCTACCGAGGCGCGGAGTATGTCTATTAGCCGCTGCC 494
QY 421 AATGCTAGCGCGCGCGGCGGAGCGCAACAGCCAGGCGCGACACCTTCTTCAACGCGCG 480
Db 495 AATGCTAGCGCGCGCGGCGGAGCGCAACAGCCAGGCGCGACACCTTCTTCAACGCGCG 554
QY 481 GGTGTTCCGGGGGCGCTTCCGGAGACCCACCTCTCGCTGCGGGGTGGCCCTCGGGCTGG 540
Db 555 GGTGTTCCGGGGGCGCTTCCGGAGACCCACCTCTCGCTGCGGGGTGGCCCTCGGGCTGG 614
QY 541 AACCCCGCATCTTACGGGCGCTTGAGCCCTTACAAGCCGCGCGGCGGGCTTCCGGGAG 600
Db 615 AACCCCGCATCTTACGGGCGCTTGAGCCCTTACAAGCCGCGCGGCGGGCTTCCGGGAG 674
QY 601 AGTCTAGCGCGCGGAGCTTGGGGCGCGCAAGCGTTTTCGTCTATCCCGCGGTACGTG 660
Db 675 AGTCTAGCGCGCGGAGCTTGGGGCGCGCAAGCGTTTTCGTCTATCCCGCGGTACGTG 734
QY 661 GAGACGCTGGTGGTCCGGAGAGTCAATGTGTCAGTTTCCAGCGCGGCGGACCTTGAACAT 720
Db 735 GAGACGCTGGTGGTCCGGAGAGTCAATGTGTCAGTTTCCAGCGCGGCGGACCTTGAACAT 794
QY 721 TATCTGTGACGCTGTGCGCAACGCGCGGAGCTTACCAGCCATCCAGCATCTCTCAAC 780
Db 795 TATCTGTGACGCTGTGCGCAACGCGCGGAGCTTACCAGCCATCCAGCATCTCTCAAC 854
QY 781 CCCATCAACATCTGTGGTCAAGGTGCTGCTTCTTAGAGATGCTGACTCCGGGCGCAAG 840
Db 855 CCCATCAACATCTGTGGTCAAGGTGCTGCTTCTTAGAGATGCTGACTCCGGGCGCAAG 914
QY 841 GTCACCGCAATGGGGCGCTGACGCTGGCAACTTCTGTGCTGGCAGAAAGCTGAAC 900
Db 915 GTCACCGCAATGGGGCGCTGACGCTGGCAACTTCTGTGCTGGCAGAAAGCTGAAC 974
QY 901 AAGTGAGTGACAGACAGCCCGAGTACTGGGACACTGCCATCTCTTCAACAGCAGGAC 960
Db 975 AAGTGAGTGACAGACAGCCCGAGTACTGGGACACTGCCATCTCTTCAACAGCAGGAC 1034
QY 961 CTGTGTGGAGCACACCTGTGACACCTGGGCTGCTGATGCTGCTGCTGCTGCTGCTGCTG 1020
Db 1035 CTGTGTGGAGCACACCTGTGACACCTGGGCTGCTGATGCTGCTGCTGCTGCTGCTGCTG 1094
QY 1021 CCCAAGAGAAGCTGCTGCTGCTTATTTGAGCAGATGGGCTTCCATCAGCCTTCAACACCTGCC 1080
Db 1095 CCCAAGAGAAGCTGCTGCTGCTTATTTGAGCAGATGGGCTTCCATCAGCCTTCAACACCTGCC 1154
QY 1081 CACGAGCTGGGCGAGTGTTCACATGCCCGCATGACATGAAAGTCTGTGAGAGGTG 1140
Db 1155 CACGAGCTGGGCGAGTGTTCACATGCCCGCATGACATGAAAGTCTGTGAGAGGTG 1214
QY 1141 TTTGGGAAGCTCCCGAGCAACACATGATGTCGCCGACCTTCATCCAGATCGACAGTGGTGC 1200
Db 1215 TTTGGGAAGCTCCCGAGCAACACATGATGTCGCCGACCTTCATCCAGATCGACAGTGGTGC 1274
QY 1201 AACCCCTGGTACGCTGAGTGTGCCATCATCCGACTTCTTGGAGACAGGCGCAGGT 1260
Db 1275 AACCCCTGGTACGCTGAGTGTGCCATCATCCGACTTCTTGGAGACAGGCGCAGGT 1334
QY 1261 GACTGCTCTTGGAGCAACCGACAGCCCATCTCCCTGCGCGGAGTATGCGCGGGCGCC 1320
Db 1335 GACTGCTCTTGGAGCAACCGACAGCCCATCTCCCTGCGCGGAGTATGCGCGGGCGCC 1394
QY 1321 AGCTACACCTTGAAGCAGCAGTGGAGCTGGCTTTTGGCTGGGCTCCAAAGCCCTGTCTCT 1380
Db 1395 AGCTACACCTTGAAGCAGCAGTGGAGCTGGCTTTTGGCTGGGCTCCAAAGCCCTGTCTCT 1454
```

```
QY 1381 TACATGAGTACTGACACCAAGCTGTGGTGCACCGGGAAGGCAAGGACAGATGTTGTC 1440
|||||
Db 1455 TACATGAGTACTGACACCAAGCTGTGGTGCACCGGGAAGGCAAGGACAGATGTTGTC 1514
|||||
QY 1441 CAGACCCGCACTTCCCTGGGCGGATGACACAGCTGTGGGAGGGCAAGCTCTGCCTC 1500
|||||
Db 1515 CAGACCCGCACTTCCCTGGGCGGATGACACAGCTGTGGGAGGGCAAGCTCTGCCTC 1574
|||||
QY 1501 AAGGGGCTGGTGGAGAGACACACCTCAACAGCAGAGGCTGATGTTCTTGGGCC 1560
|||||
Db 1575 AAGGGGCTGGTGGAGAGACACACCTCAACAGCAGAGGCTGATGTTCTTGGGCC 1634
|||||
QY 1561 AATGGGATCCCTATATGGCCCTCTGCGCACATGTGTGGGGGCTGCAGCTGGCCAGG 1620
|||||
Db 1635 AATGGGATCCCTATATGGCCCTCTGCGCACATGTGTGGGGGCTGCAGCTGGCCAGG 1694
|||||
QY 1621 AGCAGTGACCAACCCACCCCTGCCAACGGGGGCAAGTACTGAGGGAGTGAGGGTG 1680
|||||
Db 1695 AGCAGTGACCAACCCACCCCTGCCAACGGGGGCAAGTACTGAGGGAGTGAGGGTG 1754
|||||
QY 1681 AATACCGATCTTGCACCTGGAGCCCTGCCAGCTCAGCCCTCCGGAAGAGCTTCCGG 1740
|||||
Db 1755 AATACCGATCTTGCACCTGGAGCCCTGCCAGCTCAGCCCTCCGGAAGAGCTTCCGG 1814
|||||
QY 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGACCAACCGGCTCACTCTGCC 1800
|||||
Db 1815 GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGACCAACCGGCTCACTCTGCC 1874
|||||
QY 1801 GTGGCATGGTGCCCAAGTACTCCGGCGTGTCTCCCGGGACAAGTGAAGCTCATCTGC 1860
|||||
Db 1875 GTGGCATGGTGCCCAAGTACTCCGGCGTGTCTCCCGGGACAAGTGAAGCTCATCTGC 1934
|||||
QY 1861 CGAGCAATGGCACTGGCTACTTCTATGTGCTGGCACCCAA---GGTGGTGGACGGCAG 1917
|||||
Db 1935 CGAGCAATGGCACTGGCTACTTCTATGTGCTGGCACCCAAAGGTGTGTGGACGGCAG 1994
|||||
QY 1918 CTGTGCTCTCCTGACTCCACCTCCGCTGTGTGTCACCAAGGCAAGTGCATCAAGGCTGCTGT 1977
|||||
Db 1995 CTGTGCTCTCCTGACTCCACCTCCGCTGTGTGTCACCAAGGCAAGTGCATCAAGGCTGCTGT 2054
|||||
QY 1978 GATGGGAACCTGGCTCCCAAGAGATTCGACCAAGTGTGGGTGTGGGGAGACAAT 2037
|||||
Db 2055 GATGGGAACCTGGCTCCCAAGAGATTCGACCAAGTGTGGGTGTGGGGAGACAAT 2114
|||||
QY 2038 AAGAGCTGCAAGAGGTGACTGGACCTTTCACCAAGCCCATGCATGGCTACAATTTCTGTG 2097
|||||
Db 2115 AAGAGCTGCAAGAGGTGACTGGACCTTTCACCAAGCCCATGCATGGCTACAATTTCTGTG 2174
|||||
QY 2098 GTGGCCATCCCGCAGCGCCTCAAGCATCGACATCGCAGCGCGGTTACAAAGGGCTG 2157
|||||
Db 2175 GTGGCCATCCCGCAGCGCCTCAAGCATCGACATCGCAGCGCGGTTACAAAGGGCTG 2234
|||||
QY 2158 ATCGGGATGACAACTACCTGGCTCTGAAGACAGCAAGGCAAGTACCTGTCTCAACGGG 2217
|||||
Db 2235 ATCGGGATGACAACTACCTGGCTCTGAAGACAGCAAGGCAAGTACCTGTCTCAACGGG 2294
|||||
QY 2218 CATTTCTGTGTGCGGCTGTGGAGCGGACCTGTGTGGAAGGGCAAGTCTGTCTCGGCTAC 2277
|||||
Db 2295 CATTTCTGTGTGCGGCTGTGGAGCGGACCTGTGTGGAAGGGCAAGTCTGTCTCGGCTAC 2354
|||||
QY 2278 AGCGGCACGGGCACAGGGTGTGAGAGCTTCAGGCTTCCGGGCCCATCTCTGGAGCGCGTG 2337
|||||
Db 2355 AGCGGCACGGGCACAGGGTGTGAGAGCTTCAGGCTTCCGGGCCCATCTCTGGAGCGCGTG 2414
|||||
QY 2338 ACCTGTGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTTCTAT 2397
|||||
Db 2415 ACCGTGAGGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTTCTAT 2474
|||||
QY 2398 CTGCCCCAAGAGCCTCGGGAGGACAAAGTCTCTCATCCCAAGGACCCCGG---GGACCC 2454
|||||
Db 2475 CTGCCCCAAGAGCCTCGGGAGGACAAAGTCTCTCATCCCGCCAGCCCGCGGGAGGACCC 2534
|||||
QY 2455 TCTGTCTTGACACACAGCTCTCTCAGCCCTCTCCACCAAGGCTGGAGCGCGGACGACAG 2514
|||||
```

```
Db 2535 TCTGTCTTGACACACAGCTCTCAGCCTCTCCAAACAGGTGGAGCAGCGGACGACAG 2594
|||||
QY 2515 CCCCCTCAGCTCGGGTGGCTGGCAGCTGGGGCGCTGCTCCGCGAGCTCGCGCAGTGGC 2574
|||||
Db 2595 CCCCCTCAGCTCGGGTGGCTGGCAGCTGGGGCGCTGCTCCGCGAGCTCGCGCAGTGGC 2654
|||||
QY 2575 CTGAGAAGCGGGCGGTGGAATGCCGGGGTCTCCCGGGCAGCGCAGCGTCCCTGCTGT 2634
|||||
Db 2655 CTGAGAAGCGGGCGGTGGAATGCCGGGGTCTCCCGGGCAGCGCAGCGTCCCTGCTGT 2714
|||||
QY 2635 GATCAGCCCATCGGGCGGTGGAGACACAGCCCTGGCGGGGAGCCCTGCCACCTGGGAG 2694
|||||
Db 2715 GATCAGCCCATCGGGCGGTGGAGACACAGCCCTGGCGGGGAGCCCTGCCACCTGGGAG 2774
|||||
QY 2695 CTGACGCGCTGCTCACCCTGCTCCAAGAGCTGCGCGCGGGGATTTTCAGAGCGCTCACTC 2754
|||||
Db 2775 CTGACGCGCTGCTCACCCTGCTCCAAGAGCTGCGCGCGGGATTTTCAGAGCGCTCACTC 2834
|||||
QY 2755 AAGTGTGTGGGCCACGAGCGCGCTGCTGGCCCGGGACACAGTGCACCTTGCACCGCAAG 2814
|||||
Db 2835 AAGTGTGTGGGCCACGAGCGCGCTGCTGGCCCGGGACACAGTGCACCTTGCACCGCAAG 2894
|||||
QY 2815 CCCCAGGAGCTGGACTCTGCTGCTCTGAGCGCGTGC 2850
|||||
Db 2895 CCCCAGGAGCTGGACTCTGCTGCTCTGAGCGCGTGC 2930
|||||

RESULT 4
AP002986 157963 bp DNA linear PRI 08-MAR-2002
LOCUS Homo sapiens genomic DNA, chromosome 11q, clone:RP11-121M22,
DEFINITION complete sequence.
ACCESSION AP002986
VERSION AP002986.2 GI:19263031
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-121M22.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in database (2000)
REFERENCE 2 (bases 1 to 157963)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan
1-7-22 Suehiro-chou,Tsuri-mi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Mar 7, 2002 this sequence version replaced gi:11559301.
FEATURES
Location/Qualifiers
1..157963
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-121M22"

BASE COUNT 39742 a 36722 c 37247 g 44252 t
ORIGIN

Query Match 33.6%; Score 957.2; DB 9; Length 157963;
Best Local Similarity 98.2%; Pred. No. 3.8e-161;
Matches 968; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 ATGCTTTTGTGGGCACTCAACCTGCTTTCGCGGGGCAACCGCTTGGAGGCTTTGAG 60
|||||
Db 134115 ATGCTTCTGCTGGGCACTCAACCTGCTTTCGCGGGGCAACCGCTTGGAGGCTTCTGAG 134174
|||||
```

```

QY 61 CCAGAGCGGGAGGTAGTCGTTCCCATCCGACTGACCGGACATTAAGCGCGCCGCTAC 120
|
|
|
Db 134175 CCAGAGCGGGAGGTAGTCGTTCCCATCCGACTGACCGGACATTAAGCGCGCCGCTAC 134234
|
|
|
QY 121 TACTGGCGGGTCCGAGGACTCCGGGATACGGGACTCATTTTTCAGATCACAGCATTT 180
|
|
|
Db 134235 TACTGGCGGGTCCGAGGACTCCGGGATACGGGACTCATTTTTCAGATCACAGCATTT 134294
|
|
|
QY 181 CAGGAGGACTTTTACCTACACCTGACGCGGATGCTCAGTTCTTGCTCCCGCTTCTCC 240
|
|
|
Db 134295 CAGGAGGACTTTTACCTACACCTGACGCGGATGCTCAGTTCTTGCTCCCGCTTCTCC 134354
|
|
|
QY 241 ACTGAGCATCTGGCGGTCCTCCAGGGGCTCACGGGGCTCTTTCAGACCTGCGACGCG 300
|
|
|
Db 134355 ACTGAGCATCTGGCGGTCCTCCAGGGGCTCACGGGGCTCTTTCAGACCTGCGACGCG 134414
|
|
|
QY 301 TGCTTCTATTCTGGGACGTGACGCGGAGCGGAGCTGCTGCTCTGAGCCTGTGCG 360
|
|
|
Db 134415 TGCTTCTATTCTGGGACGTGACGCGGAGCGGAGCTGCTGCTCTGAGCCTGTGCG 134474
|
|
|
QY 361 GGGGGGCTCCGCGAGGCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCGCTGCC 420
|
|
|
Db 134475 GGGGGGCTCCGCGAGGCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCGCTGCC 134534
|
|
|
QY 421 AATGCTAGCGCGCGGGGCGGAGCGACAGCGAGGCGCGACACTTCTCCAGCGCGCG 480
|
|
|
Db 134535 AATGCTAGCGCGCGGGGCGGAGCGACAGCGAGGCGCGACACTTCTCCAGCGCGCG 134594
|
|
|
QY 481 GGTGTTCCGGGGGGGCTTCCGGAGACCCACTCTCGCTGGGGGTGCGCTCGGGCTGG 540
|
|
|
Db 134595 GGTGTTCCGGGGGGGCTTCCGGAGACCCACTCTCGCTGGGGGTGCGCTCGGGCTGG 134654
|
|
|
QY 541 AACCCCGCCATCTACGGGCGCTTGGACCTTACAAGCGCGGCGGGCTTCCGGGAG 600
|
|
|
Db 134655 AACCCCGCCATCTACGGGCGCTTGGACCTTACAAGCGCGGCGGGCTTCCGGGAG 134714
|
|
|
QY 601 AGTCGTAGCGCGCGAGGTCTGGGCGCGGACAGCGTTTCTGTCTATCCCGCGGTACG 660
|
|
|
Db 134715 AGTCGTAGCGCGCGAGGTCTGGGCGCGGACAGCGTTTCTGTCTATCCCGCGGTACG 134774
|
|
|
QY 661 GAGACGCTGTGTGTCGGGACGAGTCAATGGTCAAGTTCACGGGCGGAGCTGGAACAT 720
|
|
|
Db 134775 GAGACGCTGTGTGTCGGGACGAGTCAATGGTCAAGTTCACGGGCGGAGCTGGAACAT 134834
|
|
|
QY 721 TATCTGCTGACGCTGTGGGACGCGGCGGAGCTTACCGCCATCCAGCATCCTCAAC 780
|
|
|
Db 134835 TATCTGCTGACGCTGTGGGACGCGGCGGAGCTTACCGCCATCCAGCATCCTCAAC 134894
|
|
|
QY 781 CCATCAACATCTGTTGTCAGGTGCTTCTTAGAGATCGTACCTCCGGGCGGCGAAG 840
|
|
|
Db 134895 CCATCAACATCTGTTGTCAGGTGCTTCTTAGAGATCGTACCTCCGGGCGGCGAAG 134954
|
|
|
QY 841 GTACCGGCAATGCGGCGCTGAGCTGCGCAACTTCTGTGCTGCGAGAAAGCTGAAC 900
|
|
|
Db 134955 GTACCGGCAATGCGGCGCTGAGCTGCGCAACTTCTGTGCTGCGAGAAAGCTGAAC 135014
|
|
|
QY 901 AAGTGAAGTACAGCAACCGGAGTACTGGGACACTGCGCATCCTTCCACGAGGACGAG 960
|
|
|
Db 135015 AAGTGAAGTACAGCAACCGGAGTACTGGGACACTGCGCATCCTTCCACGAGGACGAG 135074
|
|
|
QY 961 CTGTGTGGAGCCACCATGTCAC 986
|
|
|
Db 135075 AGTTGATGCGCGTCACTTTGACACC 135100
|
|
|

```

```

RESULT 5
AC025130/c
LOCUS AC025130 172905 bp DNA linear HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 11 clone RP11-211H6 map 11, WORKING DRAFT
SEQUENCE, 16 unordered pieces.
ACCESSION AC025130
VERSION AC025130.2 GI:7387384
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```

```

SOURCE
ORGANISM Homo sapiens
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 11, clone RP11-211H6
JOURNAL Unpublished
REFERENCE
AUTHORS 2 (bases 1 to 172905)
1 (bases 1 to 172905)

```

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burckett, G., Campoliano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisanli, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 172905)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burckett, G., Campoliano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisanli, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 3, 2000 this sequence version replaced gi:7158941.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7808
Center clone name: 211_H6

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731
 Consensus quality: 161899 bases at least Q40
 Consensus quality: 167402 bases at least Q30
 Consensus quality: 169801 bases at least Q20
 Insert size: 177000; agarose-fp
 Insert size: 171405; sum-of-contigs
 Quality coverage: 4.1 in Q20 bases; agarose-fp
 Quality coverage: 4.2 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 16 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 21: contig of 21 bp in length
2 121: gap of 100 bp
22 1343: contig of 1222 bp in length
1344 1443: gap of 100 bp
1444 2636: contig of 1193 bp in length
2637 2736: gap of 100 bp
2737 4052: contig of 1316 bp in length
4053 4152: gap of 100 bp
4153 8044: contig of 3892 bp in length
8045 8144: gap of 100 bp
8145 14949: contig of 6805 bp in length
14950 15049: gap of 100 bp
15050 23535: contig of 8486 bp in length
23536 23635: gap of 100 bp
23636 30261: contig of 6626 bp in length
30262 30361: gap of 100 bp
30362 40941: contig of 10580 bp in length
40942 41041: gap of 100 bp
41042 54893: contig of 13852 bp in length
54894 54993: gap of 100 bp
54994 68888: contig of 13895 bp in length
68889 68988: gap of 100 bp
68989 83702: contig of 14714 bp in length
83703 83802: gap of 100 bp
83803 98393: contig of 14591 bp in length
98394 98493: gap of 100 bp
98494 117356: contig of 18863 bp in length
117357 117456: gap of 100 bp
117457 140589: contig of 23133 bp in length
140590 140689: gap of 100 bp
140690 172905: contig of 32216 bp in length.

```

FEATURES

source

```

1. .172905
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="rp11-211H6"
/clone_lib="RPCI-11 Human Male BAC"
1. .21
/note="assembly_fragment"
clone_end:T7
vector_side:right

```

misc_feature

```

122. .1343
/note="assembly_fragment"
1444. .2636
/note="assembly_fragment"
2737. .4052
/note="assembly_fragment"
4153. .8044
/note="assembly_fragment"
8145. .14949
/note="assembly_fragment"
15050. .23535
/note="assembly_fragment"
23636. .30261

```

```

/note="assembly_fragment"
clone_end:SP6
vector_side:right"
30362. .40941
/note="assembly_fragment"
41042. .54893
/note="assembly_fragment"
54994. .68888
/note="assembly_fragment"
68989. .83702
/note="assembly_fragment"
83803. .98393
/note="assembly_fragment"
98494. .117356
/note="assembly_fragment"
117457. .140589
/note="assembly_fragment"
140690. .172905
/note="assembly_fragment"
BASE COUNT 44360 a 41333 c 39516 g 46194 t 1502 others
ORIGIN

Query Match 33.6%; Score 957.2; DB 2; Length 172905;
Best Local Similarity 98.2%; Pred. No. 3.8e-161; Indels 0; Gaps 0;
Matches 968; Conservative 0; Mismatches 18;

Qy 1 ATGCTTTTGTGGCGATCCTAACCTTGCTTTGCGCGGCGGAACCGCTGAGGCTTTGAG 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 79583 ATGCTTCTGTGGCGATCCTAACCTTGCTTTGCGCGGCGGAACCGCTGAGGCTT 79524

Qy 61 CCAGAGCGGGAGGTAGTCTGTTCCCATCGACTGGACCGCGGACATTAACGGCGCGGTAC 120
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 79523 CCAGAGCGGGAGGTAGTCTGTTCCCATCGACTGGACCGCGGACATTAACGGCGCGGTAC 79464

Qy 121 TACTGGCGGGGTCCCGAGGACTCCGGGGATCCGGGACTCAATTTTCAGATCACAGCAATTT 180
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 79463 TACTGGCGGGGTCCCGAGGACTCCGGGGATCCGGGACTCAATTTTCAGATCACAGCAATTT 79404

Qy 181 CAGGAGGACTTTTACTTACACCTGACGCGGAGTGTCTAGTCTTGGCTCCGCGCTTCTCC 240
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 79403 CAGGAGGACTTTTACTTACACCTGACGCGGAGTGTCTAGTCTTGGCTCCGCGCTTCTCC 79344

Qy 241 ACTGAGCATCTGGCGTCCCTCCAGGGGTCCAGCGGGGTCTTCAGACTGCGAGCGC 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 79343 ACTGAGCATCTGGCGTCCCTCCAGGGGTCCAGCGGGGTCTTCAGACTGCGAGCGC 79284

Qy 301 TGCCTTCTATTTCTGGGAGCTGAACCGCGGACTCGTTGCTGCTGTGTAGCCTGTGC 360
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 79283 TGCCTTCTATTTCTGGGAGCTGAACCGCGGACTCGTTGCTGCTGTGTAGCCTGTGC 79224

Qy 361 GGGGGGCTCCGCGGAGCTTTGGCTACCGAGCGCGGAGTATGTTCATTTAGCCCGTGC 420
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 79223 GGGGGGCTCCGCGGAGCTTTGGCTACCGAGCGCGGAGTATGTTCATTTAGCCCGTGC 79164

Qy 421 AATGCTAGCGCGCGCGCGCAGCGCAACAGCAGGCGGACACCTTCTCCAGCGCGCG 480
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 79163 AATGCTAGCGCGCGCGCGCAGCGCAACAGCAGGCGGACACCTTCTCCAGCGCGCG 79104

Qy 481 GGTGTTCCGGCGGGCGCTTCCGGAGACCCCACTCTCGCTGCGGGGTGGCTTCGGGCTGG 540
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 79103 GGTGTTCCGGCGGGCGCTTCCGGAGACCCCACTCTCGCTGCGGGGTGGCTTCGGGCTGG 79044

Qy 541 AACCCCGCATCTCTACGGGCGCTTGACCCCTTACAGCGCGCGCGCGCGCTTCGCGGAG 600
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 79043 AACCCCGCATCTCTACGGGCGCTTGACCCCTTACAGCGCGCGCGCGCGCTTCGCGGAG 78984

Qy 601 AGTCGTAGCGCGCGAGGTCTGGGCGCGCAAGCGCTTTCGTCTATATCCCGCGGTACGTG 660
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 78983 AGTCGTAGCGCGCGAGGTCTGGGCGCGCAAGCGCTTTCGTCTATATCCCGCGGTACGTG 78924

Qy 661 GAGACGCTGGTGGTCCGCGACGAGTCAATGTTCAGTTCACGCGCGCGGACCTTGAACAT 720
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 78923 GAGACGCTGGTGGTCCGCGACGAGTCAATGTTCAGTTCACGCGCGCGGACCTTGAACAT 78864

```


QY 721 TATCTGCTGACGCTGTGCAACGGCGCGGACTTACCGCCATCCAGCATCCTCAAC 780
 |||||
 Db 78663 TATCTGCTGACGCTGTGCAACGGCGCGGACTTACCGCCATCCAGCATCCTCAAC 78804
 |||||
 QY 781 CCATCAACATCGTTGTGGTCAAGGTGCTCTTCTTAGAGATCGTGACTCCGGGCCAAG 840
 |||||
 Db 78803 CCCATCAACATCGTTGTGGTCAAGGTGCTCTTCTTAGAGATCGTGACTCCGGGCCAAG 78744
 |||||
 QY 841 GTACCCGGCAATCGGCGCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAGCTGAAC 900
 |||||
 Db 78743 GTACCCGGCAATCGGCGCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAGCTGAAC 78684
 |||||
 QY 901 AAGTGTAGTGACAGACCCCGAGTACTGGACACTGCCATCCTCTTACCAGGAGGAC 960
 |||||
 Db 78683 AAGTGTAGTGACAGACCCCGAGTACTGGACACTGCCATCCTCTTACCAGGAGGAG 78624
 |||||
 QY 961 CTGTGTGAGGCCACCACTGTGACAC 986
 |||||
 Db 78623 AGTTGATCTGCGTCACTTTTGACACC 78598
 |||||

RESULT 6
 AC023429
 LOCUS Homo sapiens chromosome 11 clone RP11-121M22, WORKING DRAFT
 DEFINITION
 AC023429
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 170682)
 AUTHORS Abola, A.P., Bruno, D., Conn, L., Della Rosa, M., Faulkner, D.,
 Federspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,
 Mao, J., Komp, C., Kottler, S., Lam, B., Marathe, R., Miranda, M.,
 Morehouse, A.J., Nguyen, M., Oefner, P., Palm, C.J., Ramirez, D.,
 Southwick, A.M., Webb, C., Wilhelmy, J., Yu, S. and Davis, R.W.
 Unpublished
 REFERENCE 2 (bases 1 to 170682)
 AUTHORS Bruno, D., Conn, L., Della Rosa, M., Faulkner, D., Federspiel, N.,
 Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R.,
 Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J.,
 Yu, S. and Davis, R.W.
 Direct Submission
 TITLE Submitted (14-FEB-2000) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 COMMENT On Apr 10, 2001 this sequence version replaced gi:13562078.
 ----- Genome Center
 Center: Stanford DNA Sequencing and Technology Development
 Center
 Center code: SDSTDC
 Web site: <http://sequence-www.stanford.edu/group/human/>
 Contact: hum-info@sequence.stanford.edu
 ----- Project Information
 Center project name: 837
 Center clone name: RP11-121M22
 ----- Summary Statistics
 Sequencing Vector: M13mp18; X02513
 Chemistry: Dye-primer; 12% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 165770 bases at least Q40
 Consensus quality: 167255 bases at least Q30
 Consensus quality: 167918 bases at least Q20
 Insert size: 172423; agarose-fp
 Insert size: 170082; sum-of-contigs
 Quality coverage: 9.0x in Q20 bases; agarose-fp
 Quality coverage: 9.1x in Q20 bases; sum-of-contigs.
 * NOTE: This is a 'working draft' sequence. It currently

* consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * * 1856: contig of 1856 bp in length
 * 1857: gap of unknown length
 * 1957: contig of 5109 bp in length.
 * 7065: gap of unknown length
 * 7066: contig of 13924 bp in length.
 * 21089: contig of 12956 bp in length
 * 21090: gap of unknown length
 * 21190: contig of 12956 bp in length
 * 34145: gap of unknown length
 * 34245: contig of 14317 bp in length
 * 34246: gap of unknown length
 * 48562: contig of 14317 bp in length
 * 48563: gap of unknown length
 * 109179: contig of 60517 bp in length
 * 109279: gap of unknown length
 * 109180: contig of 61403 bp in length.
 * 170682: contig of 61403 bp in length.

FEATURES

Source

Location/Qualifiers

1. 170682
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /clone="RP11-121M22"
 /clone_lib="RPC1 human BAC library 11"

1. 1856
 /note="assembly_name:Contig25"

1957. 7065
 /note="assembly_name:Contig26"

7166. 21089
 /note="assembly_name:Contig27"

21190. 34145
 /note="assembly_name:Contig28"

34246. 48562
 /note="assembly_name:Contig29"

48663. 109179
 /note="assembly_name:Contig30"

109280. 170682
 /note="assembly_name:Contig31"

BASE COUNT 43947 a 39732 c 39514 g 46884 t 605 others
 ORIGIN

Query Match

Best Local Similarity 31.3%; Score 893; DB 2: Length 170692;

Matches 948; Conservative 0; Mismatches 36; Indels 4; Gaps 3;

QY 1 ATGCTTTTCTGGGCGATCTAACCTTTCGCGGGCGAAACGCTGGAGGCTTTGAG 60
 |||||
 Db 98460 ATGCTTTTCTGGGCGATCTAACCTTTCGCGGGCGAAACGCTGGAGGCTTTGAG 98519
 |||||
 QY 61 CCAG-AGCGGGAGGTA-GTCGTTCCCATCCGACTGGACCCCGGACATTAACGCGGCCCT 118
 |||||
 Db 98520 CCAGAGCGGGAGGTA-GTCGTTCCCATCCGACTGGACCCCGGACATTAACGCTTTCCGT 98579
 |||||
 QY 119 ACTACTGGCGGGGTCGCCGAGGACTCCGGGATCAGGAGCTCATTTTTCAGATCAAGCAT 178
 |||||
 Db 98580 TTATCTGGCGGGTACC--GAGAGCCCGGGGTTCAGGACTCATTTTTCAGATCAAGCAT 98637
 |||||
 QY 179 TTCAGGAGGACTTTTACCTACACCTGACGCGGGATGCTCAGTTCTTGCTGCTCCCTTCT 238
 |||||
 Db 98638 TTCAGGAGGACTTTTACCTACACCTGACGCGGGATGCTCAGTTCTTGCTGCTCCCTTCT 98697
 |||||
 QY 239 CCACGTAGCATCTGGGCGTCCCTCCAGGGGCTACCGGGGCTCTTACAGACCTGCGAC 298
 |||||
 Db 98698 CCACGTAGCATCTGGGCGTCCCTCCAGGGGCTACCGGGGCTCTTACAGACCTGCGAC 98757
 |||||
 QY 299 GCTGCTTCTATCTGGGAGCTGAACGCGGAGCTCGTTCGCTGCTGAGCCTGT 358
 |||||
 Db 98758 GCTGCTTCTATCTGGGAGCTGAACGCGGAGCTCGTTCGCTGCTGAGCCTGT 98817
 |||||

```
QY 359 GCGGGGGCTCCCGGAGCCTTTGGCTACCGAGCGCCGAGATGTCATTTAGCCGCTGC 418
|||||
Db 98818 GCGGGGGCTCCCGGAGCCTTTGGCTACCGAGCGCCGAGATGTCATTTAGCCGCTGC 98877
QY 419 CCAATGCTAGCGCGCGCGGCGGACGGAACAGCGCGGCGACACCTTCTCCAGCGCC 478
|||||
Db 98878 CCAATGCTAGCGCGCGGCGGCGGACGGAACAGCGCGGCGACACCTTCTCCAGCGCC 98937
QY 479 GGGGTGTTCCGGCGCGGCTTCGGAGACCCACCTCTCGCTGCGGGGCGCTCGGGCT 538
|||||
Db 98938 GGGGTGTTCCGGCGCGGCTTCGGAGACCCACCTCTCGCTGCGGGGCGCTCGGGCT 98997
QY 539 GGAACCCCGCATCTTACGCGGCGCTTGAGACCTTACAGCGCGGCGGCGGCTTCGGGG 598
|||||
Db 98998 GGAACCCCGCATCTTACGCGGCGCTTGAGACCTTACAGCGCGGCGGCGGCTTCGGGG 99057
QY 599 AGAGTCGTAGCGCGCGGCGGCTTGAGACCTTACAGCGCGGCGGCGGCTTCGGGG 658
|||||
Db 99058 AGAGTCGTAGCGCGCGGCGGCTTGAGACCTTACAGCGCGGCGGCGGCTTCGGGG 99117
QY 659 TGGAGACGCTGTTGTCGCGGAGCTCAATGTCAGTTCACGCGCGGCGGCTTCGGGG 718
|||||
Db 99118 TGGAGACGCTGTTGTCGCGGAGCTCAATGTCAGTTCACGCGCGGCGGCTTCGGGG 99177
QY 719 ATTATCTGTCAGCTGTCGCGGAGCGGCGGCGGCTTACCGGCGGCGGCTTCGGGG 778
|||||
Db 99178 ATTATCTGTCAGCTGTCGCGGAGCGGCGGCGGCTTACCGGCGGCGGCTTCGGGG 99237
QY 779 ACCCATCAACATCGTGTGTCAGGCTGCTTCTTAGAGATCGTGTACTCCGCGGCCA 838
|||||
Db 99238 ACCCATCAACATCGTGTGTCAGGCTGCTTCTTAGAGATCGTGTACTCCGCGGCCA 99297
QY 839 AGGTCACCGGCAATGCGGCGCTGACGCTCGGCACTTCTGTGCTGGCAGAGAGCTGA 898
|||||
Db 99298 AGGTCACCGGCAATGCGGCGCTGACGCTCGGCACTTCTGTGCTGGCAGAGAGCTGA 99357
QY 899 ACAAGTAGTGACAGACCGCGGAGTACTGGGACACTGCCATCTCTTCCAGCGCAGG 958
|||||
Db 99358 ACAAGTAGTGACAGACCGCGGAGTACTGGGACACTGCCATCTCTTCCAGCGCAGG 99417
QY 959 ACCTGTGTGGAGCCACCATGTGTGACAC 986
|||||
Db 99418 TGAGTTGATCGCTGCATCTTTCACCC 99445

RESULT 7
AC101990/c
LOCUS 182656 bp DNA linear HTG 21-AUG-2002
DEFINITION Mus musculus clone RP24-371J2, WORKING DRAFT SEQUENCE, 16 unordered
pieces.
ACCESSION AC101990
VERSION AC101990.2 GI:22381363
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 182656)
Birken.B., Nusbaum.C. and Lander.E.
TITLE Mus musculus, clone RP24-371J2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 182656)
Birken.B., Linton.L., Nusbaum.C., Lander.E., Ali.A., Allen.N.,
Anderson.S., Barna.N., Bastien.V., Boguslavskiy.L., Boukhgalter.B.,
Brown.A., Camarata.J., Campopiano.A., Chang.J., Chazaro.B.,
Choepel.Y., Colangelo.M., Collins.S., Collymore.A., Cook.A.,
Cooke.P., DeArellano.K., Dewar.K., Diaz.J.S., Dodge.S., Fato.S.,
Ferreira.P., Fitzhugh.W., Gage.D., Galagan.J., Gardyna.S.,
Ginde.S., Gord.S., Goyette.M., Graham.L., Grand-Pierre.N.,
Hagos.B., Hearford.A., Horton.L., Hulme.W., Iliev.I., Johnson.R.,
Jones.C., Kamat.A., Karatas.A., Kells.C., LaRocque.K.,
Lamazares.R., Landers.T., Lehoczy.J., Levine.R., Liu.G.,
MacLean.C., Macdonald.P., Major.J., Marquis.N., Matthews.C.,
McCarthy.M., McEwan.P., McKernan.K., McPheeters.R., Meldrum.J.,
Meneus.L., Mihova.T., Mlenga.V., Murphy.T., Naylor.J., Nguyen.C.,
Norbu.C., Norman.C.H., O'Connor.T., O'Donnell.P., O'Neill.D.,
Oliver.J., Peterson.K., Phunkhang.P., Pierre.N., Pollara.V.,
Raymond.C., Retta.R., Rieback.M., Riley.R., Rise.C., Rogov.P.,
Roman.J., Rosetti.M., Roy.A., Santos.R., Schauer.S., Schupback.R.,
Seaman.S., Severy.P., Spencer.B., Stange-Thomann.N., Stojanovic.N.,
Strauss.N., Subramanian.A., Talamas.J., Tesfaye.S., Theodore.J.,
Topham.K., Travers.M., Travis.N., Trigilio.J., Vassiliev.H.,
Viel.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Ye.W.J., Young.G.,
Zainoun.J., Zemke.L., Zimmer.A. and Zody.M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 182656)
Birken.B., Nusbaum.C., Lander.E., Ali.A., Allen.N., Anderson.S.,
Barna.N., Bastien.V., Bloom.T., Boguslavskiy.L., Boukhgalter.B.,
Camarata.J., Chang.J., Chazaro.B., Choepel.Y., Collymore.A.,
Cook.A., Cooke.P., DeArellano.K., Dewar.K., Diaz.J.S., Dodge.S.,
Fato.S., Ferreira.P., FitzGerald.M., Gage.D., Galagan.J.,
Gardyna.S., Gord.S., Graham.L., Grand-Pierre.N., Hagos.B.,
Horton.L., Hulme.W., Iliev.I., Johnson.R., Jones.C., Kamat.A.,
Karatas.A., Kells.C., Landers.T., Levine.R., Lindblad-Toh.K.,
Liu.G., MacLean.C., Macdonald.P., Major.J., Matthews.C.,
McCarthy.M., Meldrum.J., Meneus.L., Mihova.T., Mlenga.V.,
Murphy.T., Naylor.J., Nguyen.C., Nicol.R., Norbu.C., Norman.C.H.,
O'Connor.T., O'Donnell.P., O'Neill.D., Oliver.J., Peterson.K.,
Phunkhang.P., Pierre.N., Raymond.C., Retta.R., Rise.C., Rogov.P.,
Roman.J., Roy.A., Schauer.S., Schupback.R., Seaman.S., Severy.P.,
Smith.C., Spencer.B., Stange-Thomann.N., Stojanovic.N., Talamas.J.,
Tesfaye.S., Theodore.J., Topham.K., Travers.M., Vassiliev.H.,
Viel.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Young.G., Zainoun.J.,
Zemke.L., Zimmer.A. and Zody.M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17060766.
All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L17833
Center clone name: 371J.2
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 178030 bases at least Q40
Consensus quality: 179964 bases at least Q30
Consensus quality: 180647 bases at least Q20
Insert size: 17200; agarose-fp
Insert size: 181156; sum-of-ctnigs
Quality coverage: 8.1 in Q20 bases; agarose-fp
Quality coverage: 7.7 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved. 361: contig of 361 bp in length
* 362 461: gap of 100 bp
* 462 1538: contig of 1077 bp in length
* 1539 1638: gap of 100 bp
```

MacLean.C., Macdonald.P., Major.J., Marquis.N., Matthews.C.,
McCarthy.M., McEwan.P., McKernan.K., McPheeters.R., Meldrum.J.,
Meneus.L., Mihova.T., Mlenga.V., Murphy.T., Naylor.J., Nguyen.C.,
Norbu.C., Norman.C.H., O'Connor.T., O'Donnell.P., O'Neill.D.,
Oliver.J., Peterson.K., Phunkhang.P., Pierre.N., Pollara.V.,
Raymond.C., Retta.R., Rieback.M., Riley.R., Rise.C., Rogov.P.,
Roman.J., Rosetti.M., Roy.A., Santos.R., Schauer.S., Schupback.R.,
Seaman.S., Severy.P., Spencer.B., Stange-Thomann.N., Stojanovic.N.,
Strauss.N., Subramanian.A., Talamas.J., Tesfaye.S., Theodore.J.,
Topham.K., Travers.M., Travis.N., Trigilio.J., Vassiliev.H.,
Viel.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Ye.W.J., Young.G.,
Zainoun.J., Zemke.L., Zimmer.A. and Zody.M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Direct Submission

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 182656)

Birken.B., Nusbaum.C., Lander.E., Ali.A., Allen.N., Anderson.S.,
Barna.N., Bastien.V., Bloom.T., Boguslavskiy.L., Boukhgalter.B.,
Camarata.J., Chang.J., Chazaro.B., Choepel.Y., Collymore.A.,
Cook.A., Cooke.P., DeArellano.K., Dewar.K., Diaz.J.S., Dodge.S.,
Fato.S., Ferreira.P., FitzGerald.M., Gage.D., Galagan.J.,
Gardyna.S., Gord.S., Graham.L., Grand-Pierre.N., Hagos.B.,
Horton.L., Hulme.W., Iliev.I., Johnson.R., Jones.C., Kamat.A.,
Karatas.A., Kells.C., Landers.T., Levine.R., Lindblad-Toh.K.,
Liu.G., MacLean.C., Macdonald.P., Major.J., Matthews.C.,
McCarthy.M., Meldrum.J., Meneus.L., Mihova.T., Mlenga.V.,
Murphy.T., Naylor.J., Nguyen.C., Nicol.R., Norbu.C., Norman.C.H.,
O'Connor.T., O'Donnell.P., O'Neill.D., Oliver.J., Peterson.K.,
Phunkhang.P., Pierre.N., Raymond.C., Retta.R., Rise.C., Rogov.P.,
Roman.J., Roy.A., Schauer.S., Schupback.R., Seaman.S., Severy.P.,
Smith.C., Spencer.B., Stange-Thomann.N., Stojanovic.N., Talamas.J.,
Tesfaye.S., Theodore.J., Topham.K., Travers.M., Vassiliev.H.,
Viel.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Young.G., Zainoun.J.,
Zemke.L., Zimmer.A. and Zody.M.

Direct Submission

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Aug 21, 2002 this sequence version replaced gi:17060766.

All repeats were identified using RepeatMasker:

Snit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L17833

Center clone name: 371J.2

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 178030 bases at least Q40

Consensus quality: 179964 bases at least Q30

Consensus quality: 180647 bases at least Q20

Insert size: 17200; agarose-fp

Insert size: 181156; sum-of-ctnigs

Quality coverage: 8.1 in Q20 bases; agarose-fp

Quality coverage: 7.7 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 16 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved. 361: contig of 361 bp in length

* 362 461: gap of 100 bp

* 462 1538: contig of 1077 bp in length

* 1539 1638: gap of 100 bp

*

* 1639 3397: contig of 1759 bp in length
 * 3398 3497: gap of 100 bp
 * 3498 4947: contig of 1450 bp in length
 * 4948 5047: gap of 100 bp
 * 5048 6893: contig of 1846 bp in length
 * 6894 6993: gap of 100 bp
 * 6994 10395: contig of 3402 bp in length
 * 10396 10495: gap of 100 bp
 * 10496 12584: contig of 2089 bp in length
 * 12585 12684: gap of 100 bp
 * 12685 14977: contig of 2293 bp in length
 * 14978 15077: gap of 100 bp
 * 15078 18181: contig of 3104 bp in length
 * 18182 18281: gap of 100 bp
 * 18282 22654: contig of 4373 bp in length
 * 22655 22754: gap of 100 bp
 * 22755 31261: contig of 8507 bp in length
 * 31262 31361: gap of 100 bp
 * 31362 43587: contig of 12226 bp in length
 * 43588 43687: gap of 100 bp
 * 43688 57612: contig of 13925 bp in length
 * 57613 57712: gap of 100 bp
 * 57713 77467: contig of 19755 bp in length
 * 77468 77567: gap of 100 bp
 * 77568 108724: contig of 31157 bp in length
 * 108725 108824: gap of 100 bp
 * 108825 182656: contig of 73832 bp in length.

FEATURES

Location/Qualifiers
 1..182656
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="RPCI-24 Male Mouse BAC"
 1..361
 /note="assembly_fragment"
 462..1538
 /note="assembly_fragment"
 1639..3397
 /note="assembly_fragment"
 3498..4947
 /note="assembly_fragment"
 5048..6893
 /note="assembly_fragment"
 6994..10395
 /note="assembly_fragment"
 10496..12584
 /note="assembly_fragment"
 12685..14977
 /note="assembly_fragment"
 15078..18181
 /note="assembly_fragment"
 18282..22654
 /note="assembly_fragment"
 22755..31261
 /note="assembly_fragment"
 31362..43587
 /note="assembly_fragment"
 43688..57612
 /note="assembly_fragment"
 57713..77467
 /note="assembly_fragment"
 77568..108724
 /note="assembly_fragment"
 108825..182656
 /note="assembly_fragment"
 BASE COUNT 51468 a 41160 c 39770 g 48758 t 1500 others
 ORIGIN

Query Match 27.3%; Score 779.8; DB 2; Length 182656;
 Best Local Similarity 87.1%; Pred. No. 1.6e-129;
 Matches 856; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 1 ATGCTTTTGTGGGCATCTTAACCTGGCTTCGGCGGGCGAACCCTGGAGGCTTTGAG 60

Db 39782 ATGCTTCTACTGGGATCTCAATCTCGCTCGCGACCTGCTGGAGCTCCGAG 39723
 QY 61 CCAGAGCGGAGGTAGTCTCCATCCGACTGGACCGGACATTAAGCGGCGCGCTAC 120
 Db 39722 CCAGAGTGGAGGTGGTCTTCCAAATCCGACGGGACCAATCAATGGCGGCACAT 39663
 QY 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGACATTTTTCAGATCACACATTT 180
 Db 39662 TACAGAGGGTACGAGGACTCCGGGATCAGGCTCAATTTTTCAGATCACAGCTTT 39603
 QY 181 CAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCCGCTTCTCC 240
 Db 39602 CAGCAGGACTTTTATCTACACCTGACCGCAGATGCCAGTCTCGCTCCCGCTTCTCC 39543
 QY 241 ACTGAGATCTGGGGTCCCGCTCCAGGGGCTCACGGGGGCTCTTCAGAGCTCGGAGCG 300
 Db 39542 ACTGAGTATAGTGTCTCCGCTGCAGAGGCTCACTGGCAGCTCTAGACCTGAGCGC 39483
 QY 301 TCGTCTTATTCGGGACGTGAACCGCGGACTCGTTGCTGCTGCTGCTGCTGCTGCTGCTG 360
 Db 39482 TCGTCTTATTCGGGATGTGAACCGCGGAGCAGACTCTTTCGCTGCTGCTGCTGCTG 39423
 QY 361 GGGGGCTCCGCGGAGCCTTTGGCTACCGAGCGCGGAGTATGTCTATAGCCGCTGCC 420
 Db 39422 GGGGGTCTCCGCGAGCCTTTGGCTACCGAGGTGGAGTATGTCTATAGCTCTGCC 39363
 QY 421 AATGCTAGCGCGCGGGCGGCGAGCAACGCCAGGGGCGACACTTCTCCAGGCGCGG 480
 Db 39362 AACACCGAGCGCGGAGCGGAGCGGCTCACGCCAGGGGCGACACTTCTCCAGCGCGG 39303
 QY 481 GGTGTTCCGGGGGCGCTTCGCGAGACCCACTCTCGCTGCGGGGCTGCTCGGGCTGG 540
 Db 39302 GGTGCTCTGTAGGCGCTTCGCGAGACCCCACTCTCGTGGGGGTGGCTCGTGGCTG 39243
 QY 541 AACCCCGCATCTACGGGCGCTTGACCCCTTACAAGCGCGCGGGCGGGCTTGGGGGAG 600
 Db 39242 AACCCCGCATCTACGGGCTCTGGACCTTATAAGCCACCGCGGAGCGGCGGAG 39183
 QY 601 AGTCGTAGCGCGGAGCTCTGGGGCGCGCAAGCGTTTCGTGTCTATCCCGCGGTACGTG 660
 Db 39182 AGCCACAACCGCGGAGCTCTGGGGCGCGCAAGCGCTTCGTGTCTATACCGAGTACGTG 39123
 QY 661 GAGAGCTGGTGTGCGCGGAGGTCATGTCAAGTTCACGCGCGCGGAGCTGGAACAT 720
 Db 39122 GAGACACTGGTGTGGCGGAGGTCATGTCAAGTTCACGCGCGGAGTTCGGAACAT 39063
 QY 721 TATCTGTGACGCTGCTGGCAACCGCGGCGGAGCTCTACCGCCATCCAGCATCCTCAAC 780
 Db 39062 TATCTACTGACGCTGCTGGCCACCGCGGCGGAGCTCTACCGCCATCCAGCATCCTCAAC 39003
 QY 781 CCCATCAACATCGTGTGGTCAAGGTGCTGCTTCTTAGAGATGCTGACTCGGGGCGCAAG 840
 Db 39002 CCTATCAACATCGTGTGGTCAAGGTGCTGCTTCTTAGAGATGCTGACTCGGGGCGCAAG 38943
 QY 841 GTCACCGCAATGGCGGCGCTGACGCTGGCAACTTCTGTGCTGGCAGAGCAAGCTGAAC 900
 Db 38942 GTCACAGCAACGCGGCGCTGACTCTGGCAACTTCTGTGCTGGCAGAGCAAGCTGAAC 38883
 QY 901 AAAGTGTGACAGACAGCCCGAGTACTGGACACTGCAATCTCTTTCACAGGAGGAG 960
 Db 38882 AAAGTGTGAGGATAGACACCTTGTAGTGGGACAGCACTCTCTTTCACAGGAGGAG 38823
 QY 961 CTGTGTGGAGCCACCACTGTGA 983
 Db 38822 AGTTCACTGTACACTCTTTTGA 38800

RESULT 8
 AP003459
 LOCUS
 DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:RP11-211H5,
 complete sequence.

AP003459
VERSION AP003459.2 GI:19263045
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-211H5.
ORGANISM Homo sapiens
Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
Hattori M., Ishii K., Toyoda A., Taylor T.D., Hong-Seog P.,
Fujiyama A., Yada I., Totoki Y., Watanabe H. and Sakaki Y.
Homo sapiens genomic DNA
Published Only in Database (2001)
REFERENCE
Hattori M., Ishii K., Toyoda A., Taylor T.D., Hong-Seog P.,
Fujiyama A., Yada I., Totoki Y., Watanabe H. and Sakaki Y.
Direct Submission
Submitted (29-MAR-2001) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
On Mar 7, 2002 this sequence version replaced gi:13488920.
COMMENT
FEATURES
source
1..28000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-211H5"
BASE COUNT 6938 a 6083 c 5679 g 8300 t
ORIGIN
Query Match 27.3%; Score 778.8; DB 9; Length 28000;
Best Local Similarity 98.5%; Pred. No. 3.3e-129;
Matches 786; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Qy 2056 ACTGGACTTTCACCAAGCCATCGCTACATTTCTGTGGCCATCCCGCAGGC 2115
Db 3450 ACGGCCCCCTTCCCGCCGAGCAGTGGCTACATTTCTGTGGCCATCCCGCAGGC 3509
Qy 2116 GCCTCAAGCATCGACATCCGCGCGGGTTACAAAGGCTGTATCGGGGATGACAACTAC 2175
Db 3510 GCCTCAAGCATCGACATCCGCGCGGGTTACAAAGGCTGTATCGGGGATGACAACTAC 3569
Qy 2176 CTGGCTCTGAAGACAGCAAGCAAGTACCTGCTCAAGCGCATTTCTGTGGTTCGGCG 2235
Db 3570 CTGGCTCTGAAGACAGCAAGCAAGTACCTGCTCAAGCGCATTTCTGTGGTTCGGCG 3629
Qy 2236 GTGGAGCGGACCTGGTGTGAAGGCGAGTCTGCTGCGGTACAGCGCAGCGGCACAGCG 2295
Db 3630 GTGGAGCGGACCTGGTGTGAAGGCGAGTCTGCTGCGGTACAGCGCAGCGGCACAGCG 3689
Qy 2296 GTGGAGAGCTCGAGGCTTCCCGCCCATCTCTGAGCGCGTGTACCGGTGGAGGTCTCTCC 2355
Db 3690 GTGGAGAGCTCGAGGCTTCCCGCCCATCTCTGAGCGCGTGTACCGGTGGAGGTCTCTCC 3749
Qy 2356 GTGGGGAAGATGACACCGCCCGGGTCTGCTACTCTTCTATCTGTCCTCAAGAGCCTCGG 2415
Db 3750 GTGGGGAAGATGACACCGCCCGGGTCTGCTACTCTTCTATCTGTCCTCAAGAGCCTCGG 3809
Qy 2416 GAGGACAAAGTCTCTCATCCCAAGACCCCGGGGACCTCTGCTTGTTCACCAAGCGGTC 2475
Db 3810 GAGGACAAAGTCTCTCATCCCAAGACCCCGGGGACCTCTGCTTGTTCACCAAGCGGTC 3869
Qy 2476 CTCAGCCTCTCAACAGGTGGAGCAGCGGACGACAGCGGCCCTGTACCGCTGGGTGGCT 2535
Db 3870 CTCAGCCTCTCAACAGGTGGAGCAGCGGACGACAGCGGCCCTGTACCGCTGGGTGGCT 3929
Qy 2536 GCGAGCTGGGGCGGTGCTCCGAGCTGCGGCAAGTGGCTGTGAGAGCGGGCGGTGGAC 2595
Db 3930 GCGAGCTGGGGCGGTGCTCCGAGCTGCGGCAAGTGGCTGTGAGAGCGGGCGGTGGAC 3989
Qy 2596 TGCCGGGGCTCCCGGGGACGCGCAGCGGTGCTGCTGCTGTGATGACGCCCATCGGCCCGGTG 2655

|||||
Db 3990 TGCCGGGGCTCCCGGGGACGCGCAGCGGTGCTGCTGCTGTGATGACGCCCATCGGCCCGGTG 4049
Qy 2656 GAGACACAGCTCGGGGAGCGCTGCCCCACCTCGGGAGCTCAGCGCTGTGTACCTCGTGC 2715
Db 4050 GAGACACAGCTCGGGGAGCGCTGCCCCACCTCGGGAGCTCAGCGCTGTGTACCTCGTGC 4109
Qy 2716 TCCTAAGAGCTCGGGGAGCTTTCAGAGGCGCTCACTCAAGTGTGTGGGCGACGAGGC 2775
Db 4110 TCCTAAGAGCTCGGGGAGCTTTCAGAGGCGCTCACTCAAGTGTGTGGGCGACGAGGC 4169
Qy 2776 CGGCTGCTCGGGGAGCGCTGCAACTGTCACCGCAGCGCCAGGAGCTGAGCTCTGTC 2835
Db 4170 CGGCTGCTCGGGGAGCGCTGCAACTGTCACCGCAGCGCCAGGAGCTGAGCTCTGTC 4229
Qy 2836 GTCCTGAGGCGGTGCTGTA 2853
Db 4230 GTCCTGAGGCGGTGCTGTA 4247

AC126507 178764 bp DNA linear HTG 24-JUL-2002
Rattus norvegicus clone CH230-254N12. *** SEQUENCING IN PROGRESS
*** 49 unordered pieces.
AC126507 1 GI:21700463
HTG: HTGS_PHASE1.
Rattus norvegicus.
Rattus norvegicus.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 178764)
Muzny D.M., Adams C., Adio-Oduola B., Ali-Osman F.R., Allen C.,
Alsbrooks S.L., Amarantunga H.C., Are J.R., Ayale M., Banks T.,
Barbala J., Benton J., Bimaga K., Blankenburg K., Bonnin D.,
Bouck J., Bowle S., Brieva M., Brown E., Brown M., Bryant N.P.,
Buhay C., Burch P., Burkett C., Burrell K.L., Byrd N.C.,
Carroll F.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
Chen G., Chen R., Chen Z., Chowdhury I., Christopoulos C.,
Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,
DeLaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.H.,
Dunaway K.J., Draper H., Dugan-Rocha S., Durbin K.J.,
Earhart C., Edgar B., Edwards C.C., Elhaj C., Escotto M.,
Fallis T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,
Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,
Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J.,
Hernandez O., Hodgson A., Hogue M., Holloway C., Hollins B.,
Homs F., Howard S., Huber J., Hulyk S., Hume J., Jackson L.E.,
Jacobson B., Jia F., Johnson R., Jollivet S., Joudah S.,
Karlsson E., Kelly S., Khan U., King L., Korvan J., Kovar C.,
Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C., Lewis L.,
Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W., Louised H.,
Lozano R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
Maheshwari M., Mapua P., Martin R., Martindale A., Martinez E.,
Massey G., Mawhney E., McLeod M.P., Meador M., Mei G., Metzger M.,
Miner G., Miner Z., Mitchell T., Mohabbat K., Morgan M., Morris S.,
Moser M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N.,
Nguyen N., Nickerson E., Nwokenwo S., Ogih M., Okuwonu G.,
Oragunye N., Oviedo R., Pace A., Payton B., Peery J., Perez L.,
Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y.,
Rivers M., Rojas A., Rojubokan I., Roife M., Ruiz S., Savary G.,
Scherer S., Scott G., Shen H., Shoshitari N., Sison I.,
Sodergren E., Sonaike T., Sparks A., Stanley H., Stone H.,
Sutton A., Svatek A., Tabot P., Tamerisa A., Tamerisa K., Tang H.,
Tansey J., Taylor C., Taylor T., Telford B., Thomas R., Thomas S.,
Usmani K., Vasquez L., Vera V., Villalon D., Vinson R., Wang Q.,
Wang S., Ward-Moore S., Warren R., Washington C., Watlington S.,
Williams G., Williamson A., Wleczyk R., Wooden C., Worley K.,
Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Nelson D.,

Weinstock, G. and Gibbs, R.

TITLE

Direct Submission

REFERENCE

2 (bases 1 to 178764)

AUTHORS

Worley, K.C.

TITLE

Direct Submission

REFERENCE

Submitted (06-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

AUTHORS

Worley, K.C.

TITLE

Direct Submission

REFERENCE

Submitted (24-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GZFU

Center clone name: CH230-254N12

----- Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 124573 bases at least Q40

Consensus quality: 131850 bases at least Q30

Consensus quality: 137700 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 49 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

1 1037: contig of 1037 bp in length

* 1038: contig of unknown length

* 1138: contig of 1113 bp in length

* 2250: contig of unknown length

* 2350: contig of 1229 bp in length

* 3579: contig of unknown length

* 3680: contig of 1195 bp in length

* 4874: contig of unknown length

* 4975: contig of 1551 bp in length

* 6525: contig of unknown length

* 6626: contig of 1280 bp in length

* 7906: contig of unknown length

* 8006: contig of 1719 bp in length

* 9725: contig of unknown length

* 9825: contig of 1618 bp in length

* 11442: contig of unknown length

* 11543: contig of 1358 bp in length

* 12901: contig of unknown length

* 13001: contig of 1150 bp in length

* 14151: contig of unknown length

* 14251: contig of 1315 bp in length

* 15565: contig of unknown length

* 15666: contig of 1810 bp in length

* 17476: contig of unknown length

* 17576: contig of 1115 bp in length

* 18691: contig of unknown length

* 18791: contig of 1479 bp in length

* 20269: contig of unknown length

* 20369: contig of unknown length

* 21927: contig of 1558 bp in length

* 22027: contig of unknown length

* 22028: contig of 1971 bp in length

* 23999: contig of unknown length

* 24099: contig of 1581 bp in length

* 25680: gap of unknown length

* 25779: contig of 1649 bp in length

* 27428: contig of unknown length

* 27528: gap of unknown length

* 27529: contig of 2149 bp in length

* 29678: gap of unknown length

* 29777: contig of 1100 bp in length

* 30877: contig of unknown length

* 30978: gap of unknown length

* 32831: contig of 1854 bp in length

* 32931: gap of unknown length

* 35419: contig of 2488 bp in length

* 35519: gap of unknown length

* 37393: contig of 1874 bp in length

* 37493: gap of unknown length

* 39835: contig of 2342 bp in length

* 39935: gap of unknown length

* 42789: contig of 2854 bp in length

* 42889: gap of unknown length

* 46646: contig of 3757 bp in length

* 46746: gap of unknown length

* 49188: contig of 2442 bp in length

* 49288: gap of unknown length

* 53805: contig of 4517 bp in length

* 53905: gap of unknown length

* 56202: contig of 2297 bp in length

* 56302: gap of unknown length

* 58719: contig of 2417 bp in length

* 58819: gap of unknown length

* 62904: contig of 4085 bp in length

* 63004: gap of unknown length

* 65932: contig of 2928 bp in length

* 66032: gap of unknown length

* 70120: contig of 4088 bp in length

* 70220: gap of unknown length

* 73358: contig of 3038 bp in length

* 73559: gap of unknown length

* 77466: contig of 4108 bp in length

* 77566: gap of unknown length

* 77567: contig of 2979 bp in length

* 80545: gap of unknown length

* 80646: contig of 5342 bp in length

* 85987: gap of unknown length

* 86087: contig of 4821 bp in length

* 90908: gap of unknown length

* 91008: contig of 6910 bp in length

* 97918: gap of unknown length

* 98018: contig of 5484 bp in length

* 103502: contig of unknown length

* 103602: gap of unknown length

* 112570: contig of 8968 bp in length

* 112671: gap of unknown length

* 117924: contig of 5254 bp in length

* 118024: gap of unknown length

* 125019: contig of 6995 bp in length

* 125119: gap of unknown length

* 134087: contig of 8968 bp in length

* 134187: gap of unknown length

* 143851: contig of 9664 bp in length

* 143951: gap of unknown length

* 151699: contig of 7748 bp in length

* 151799: gap of unknown length

* 158826: contig of 7027 bp in length

* 158926: gap of unknown length

* 158927: contig of 9596 bp in length

* 168523: gap of unknown length

* 168623: contig of 10142 bp in length.

Location/Qualifiers

1. 178764

/organism="Rattus norvegicus"

/db_xref="taxon:10116"

/clone="CH230-254N12"

FEATURES

source

47487 a 39119 c 38897 g 45423 t 7838 others

BASE COUNT

ORIGIN

...AAAGTGAAGCTCATCTGCCGAGCCAATGGCACTGGCTACTTCATGTGCTGGCACCCAA 1901

Db 13 CAAGTGAAGCTAATCTGCCGAGCAATGGGACTGGCTACTTCTATGTAAGCACTAA 72
QY 1902 GGTGTGACGCGACGCTGCTCTCTGAGCTCCACCTCGTCTGTCTCAAGCAAGTG 1961
Db 73 GGTGTGAGTGTAGCTGTGCTCTGAGCTCCACCTCGTCTGTCTCAAGCAAGTG 132
QY 1962 CATCAAGCTGGCTGTGATGGGAACTCGGCTCCAAAGAGAGATTCGACAAAGTGGGGT 2021
Db 133 CATCAAGCTGGCTGGCAGCGGAATCTGGCTCCAAAGAGAAATTTGACAAATGTGGTGT 192
QY 2022 GTGTGGGGGACAAATAGAGCTCCAGAGAGTGTGAGTCTTTCACCAAGCCCAATGCA 2081
Db 193 GTGTGGTGGAGACAATAAGAGCTTAAGAGGGTGACAGGACTCTTCACCAAGCTATGCA 252
QY 2082 TGGCTACAATTTCTGCTGGCATCCCGCAGGCGCTCAAGCATCGACATCCGCCAGCG 2141
Db 253 CGGCTACAATTTCTAGTGGCATCCCTGCTGGCGCTCCAGCATTTGATATCCGTGAGCG 312
QY 2142 CGGTTACAAGGGCTGATCGGGGATGACAACTACCTGGCTCTGAAGAACCAAGCAAGCAA 2201
Db 313 TGGTTACAAGGGGCTCATTTGGAGATGACAACTACCTGGCTTTGAAGAACCAAGCAAGCAA 372
QY 2202 GTACTGCTCAAGCGCATTTCTGCTGGTGTGGGCTGGAGCGGACCTGTGTGTGAAGG 2261
Db 373 ATACTGCTCAATGGGCATTTTGGGTATCCGCTGTAGAGCGGGACCTGTGTGTAAAGG 432
QY 2262 CAGTCTGCTGGGTACAGCGCGACGCGGACAGCGGTGGAGAGCTGACAGCTTCCCGGCC 2321
Db 433 CAGTGTGCTACGCTATAGTGGCACTGTACTGCAAGTGGAGAGCTGCAGGCTTCTGACC 492
QY 2322 CATCTGGAGCGCTGACGTGGAGTCTCTCGTGGGGAGAGATGACACCGCCCGGGT 2381
Db 493 CATCTGGAGCGCACTACCGTGGAGTCTCTCGTGGGGAGAGATGACACCGCCCGGT 552
QY 2382 CCGTACTCTCTTCTATCTGCCAAAGAGCTCGGGAGCAAGTCTCTCATCCCAAGGA 2441
Db 553 GCGTATCTCTTCTACTGCTGCCAAAGAACCTCGGGAGCAAGTCTCTCATCCCAAGGA 612
QY 2442 CCCCCGGGACCCCTCTCTTGCACAAACAGCGTCTCAGCTCTCCAAACAGGTGGAGCA 2501
Db 613 CCTCGAGGTTCCCTCGTGTGGCAACAGTGTCTCAGCTTTTCCAAACAGTAGACA 672
QY 2502 GCGGAGCAGACGCGCCCTCGACGCTGGTGGCTGGCAGCTGGGGCGCTGCTCCGGAG 2561
Db 673 GCGAGATACAGACCTCTCGCGCTGGGTGGCAGGAGCTGGGGACCTTGTCTCGGTGAG 732
QY 2562 CTGCGGAGTGGCTGACAGAGCGGCGGTGGACTGCGGGGCTCGCGCGGAGCGAC 2621
Db 733 CTGTGGCAGTGGCTACAGAGCGAGCAGTGGACTGCGGGACTCCCGAGGACGCAAGG 792
QY 2622 GGTCCCTGCTGTGATGACGCCATCGGCCCTGGGAGACACAGCTGCGGGAGCCCTG 2681
Db 793 AGCTCTGCTGTGATGATAGACCATCGGCCATTGAGAGCGAGCTTGGGGAAACCTG 852
QY 2682 CCCCACCTGGAGCTCAGCGCTGGTCAACCTGCTCCAGAGCTGGGGCGGAGTTTCA 2741
Db 853 CCGAATTTGGAGCTCGGCAACTGCTCGCCCTGTCTTAAAGCTGGCGGTGGATTA 912
QY 2742 GAGCGCTCACTCAAGTGTGTGGGCCACGAGGCGGCTGCTGGCCCGGAGCACAGTGCA 2801
Db 913 GAGCGCTCCGCTCAAGTGTGTGGGCCACGAGGCGAGCTGCTGGCTCGGGACAGTGTGA 972
QY 2802 CTTGACCCGCAAGCCCGAGAGTGGACTTCTGCTGCTGAGCGCGTGTGA 2853
Db 973 CTTACGCGCAAGCCCGAGGAATTAGACTTCTGCGTCTTGAGACCTGCTGA 1024

RESULT 11
E55282 LOCUS 2670 bp DNA linear PAT 31-JAN-2002
DEFINITION Novel metalloprotease and gene of the same.
ACCESSION E55282
VERSION E55282.1 GI:18629795

KEYWORDS JP 2001008687-A/18.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2670)
AUTHORS Yamaji.N., Nishimura.K. and Sasamata.M.
TITLE Novel metalloprotease and gene of the same
JOURNAL Patent: JP 2001008687-A 18 16-JAN-2001;
YAMANOUCHI PHARMACEUT CO LTD
COMMENT OS Homo sapiens (human)
PN JP 2001008687-A/18
PD 16-JAN-2001
PF 25-JUN-1999 JP 1999180973
PR NOBORU YAMAJI, KOICHI NISHIMURA, MIHO SASAMATA
PC C12N15/09, C07K16/40, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC
C12N9/64, C12Q1/37,
PC C12N15/00, C12N5/00
CC
FH Key Location/Qualifiers
FT source 1..2670
FT /organism='Homo sapiens (human)'.
FEATURES
source
Location/Qualifiers
1..2670
/organism='Homo sapiens'
/db_xref='taxon:9606'
BASE COUNT 511 a 853 c 867 g 439 t
ORIGIN
Query Match 23.3%; Score 666; DB 6; Length 2670;
Best Local Similarity 59.2%; Pred. No. 6.6e-109;
Matches 1355; Conservative 0; Mismatches 885; Indels 48; Gaps 11;
QY 157 CTCATTTTTCAGATCACAGCATTTTCAGGAGACTTTTACCTACCTACCTGAGCGCGATGCT 216
Db 157 CTCGCGCTCCACTGTGCGCTTCGGCAAGGCTTCTGCTGCGCTGCGCGCCGACAC 216
QY 217 CAGTTCTTGGCTCCCGCTTCTCCACTGAGCATCTGGCGTGGCGTCCCTCCAGGGGCTAC 276
Db 217 AGCTTCTTGGCGCGGCTTCAAGATCGAGCGCTTCGGGGGCTCCGGCGGGCG---ACC 273
QY 277 GGGGCTCTTTCAGACTCGACGCTGCTTCTATTCTGGGAGCTGAACCGCAGCGGAC 336
Db 274 GGGGCGAGCGGGGCTGCGCGCTGCTTCTTCTCGGCACCGTGAATGGGAGCGCGAG 333
QY 337 TCGTTGCTGCTGTCAGCTGTGCGGGGCTCCCGGAGCTTTGGCTACCGAGCGCC 396
Db 334 TCGTGGCGGCGCTGAGCGCTGTGCGGCGGCTGAGCGGCTCTTCTGCTGGACGGCGAG 393
QY 397 GAGTATGTCTATTAGCCCGCTGCCCAATGCTAGCGCGCGCGGGCGCAGCGCAACAGCCAG 456
Db 394 GAGTTCAACATCCAGCCGAGCGCGGGGCTCTCTGGCTCAGCCGACCGCTGCGAG 453
QY 457 GCGGCACACCTTCTCCAGCGCGGGGTGTTCGGGCGGGCTTTCGGAGACCCCACTCT 516
Db 454 CGCTGGGCTCCCGCGGAGCGCCCGCTCCCGCGAGGACCGCGAGTGGAGGTGAGAGC 513
QY 517 CGCTCGGGGGTG-GCTCGGGGTGGAACCGCCCATCTACGGGCGCTTGAGCCCTTACAA 575
Db 514 GGAGAGGTTCAGAGCAGGAGAGAGACCAACAGAGAGACAGCGAGAGAGAGAGCCAA 573
QY 576 GCGCGGGCGGGGCTTTCGGGGAGAGTCTGAGCGGGCGCAGG-----TCTGGCGCGCC 630
Db 574 GAAGAGAGGCGAAGAGCGCTAGCGAGCGCCACCGCCCTGGGGGCCACGAGTAGGACC 633
QY 631 AAGCGTTTCTGCTCTATCCCGGGTACGTGGAGAGCGCTGTGTGGTGGCGGAGTCAATG 690
Db 634 AAGCGTTTGTCTCTGAGCGCGCTTCGTGAAGAGCGCTGTGTGGCGCGATCGCTCCATG 693
QY 691 GTCAAGTTCCAGCGCGGACCTGGAACATTTATCTCTGACGCTGCTGGCAACGCGCGG 750
Db 694 GCTGCTTCTACGGGGCGGACCTGCGAAGAACCATCTCTGAGCTTAATGTC---GTGACAGCC 753

QY 751 CGACTCTACCCGCCATCCAGCATCTCTCAACCCCATCAACATCGTTGTGTCAAGGTGCTG 810
DB 754 CGAATCTACAAGCAGCCAGCATCAAGAAATTCATCAACCTGATGTGTGTAAGTGTGCTG 813
QY 811 CTTCTTAGAGATCTGACTCGGCGCCCAAGTCTACCGGCAATCGGCCCTGACGCTGCCG 870
DB 814 ATCGTAGAGATGAAAATATGGGCCCCAGAGGTGTCCGACAAATGGGGGCTTTACACGCT 873
QY 871 AACTTCTGTGCTGGCAGAGAACTGAACAAAGTGAAGTGAACAGCAGCCCGAGTACTGG 930
DB 874 AACTTCTGCAACTGGCAGCGCTTTCAACAGCCAGCAGCCAGCCAGAGCACTAC 933
QY 931 GACACTGCTATCTTTCACAGGAGGAGCTGTGTGG-----AGCCACCACTGTGACACC 987
DB 934 GACACGGCCATCTCTCTCACCAGACAGAACTTCTGTGGCAGGAGGGGCTGTGTGACACC 993
QY 988 CTGGCATGGCTGATGTGGGTACCATGTGTGACCCCAAGAGAGCTGCTGTGTCATGTAG 1047
DB 994 CTGGGTGGGAGACATCTGGGACCATTTGTGACCCCAACAAAGCTGCTCCGTGATCGAG 1053
QY 1048 GACATGGGCTTCCATCAGCTTTCACCACTGCCACAGAGCTGGGCCACGCTGTTCACATG 1107
DB 1054 GATGAGGGCTCCAGGCGGCCACACCTTGCCCATGAATAGGACAGCTCTCAGCATG 1113
QY 1108 CCCCATGACAAATGTGAAGTCTGTGAGAGGTGTTTGGAGAGTCCGAGCCACACATG 1167
DB 1114 CCCACAGCAGACTCAAGCCCTGCACACGCTCTTCGGGCCCATGGGCAAGCACCAGTG 1173
QY 1168 ATGTCCCGGACCTTCATCAGATGAGCCGTGCCAACCCTGTGTCAGCTGCAGTGTGCC 1227
DB 1174 ATGGCAGCGTGTGCTGCTCACTGAACACAGAGCTGCGCTGGTCCCTGCAGCGCATG 1233
QY 1228 ATCATCACCGACTTCTGGAGAGCGGACGGTGAAGTCTCTCTGGACCAACCCAGCAAG 1287
DB 1234 TATCTCAGAGACTTCTGGAGCGGCGGACGAGAGTGTCTCTGGATGCCCTGTGCG 1293
QY 1288 CCCATCTCCCTGGCAGGATCTGCCGGC-----GCCAGCTACACCTGTAGCGCAGCAG 1341
DB 1294 GCGCTGCCCTCCCAAGCGCTCCGGGCGGATGCGCTGTACAGCTGGACCGCAGCAG 1353
QY 1342 TCGAGCTGGCTTTTGGCTGGCTTCAAGCCCTGTCTTACA-----TGCAGTAC 1392
DB 1354 TCGAGGAGATCTTTGGCGCGATTTCCGCCACTGTCCCAACACCTGTGCTCAGGAGTC 1413
QY 1393 TGCACCAAGCTGTGGT-----CACCGGAAGCCCAAGGAGCAGATGTGTGCCAGACCCG 1449
DB 1414 TCGCCCAAGCTTGTGTGCCACACTGTATGGGCTGAGCCCTGTGCCACAGAAATGGC 1473
QY 1450 CACTTCCCTGGGCGGATGGCAGCTGTGGGAGGSCAAGCTCTGCCTCAAGGGGC 1509
DB 1474 AGCTTGGCTGGGCTGACGCGCTGGGCGCTGGGCGCTGGGACCTCTGCTCAGAAGGCGC 1533
QY 1510 TG-----CGTGAGAGACAACTCAACAAAGCAGAGTGGATGTTCTTGGGCCAAA 1563
DB 1534 TGTCTACTGAGGAGAGTGGAGAGCCCAAGCCCTGTAGATGAGGCTGGCACCG 1593
QY 1564 TGGGATCCCTATGGCCCTCTGTCGCGACATGTGTGGGCGGTGTCAGCTGGCCAGAGG 1623
DB 1594 TGGGAGCCCTGGGAGAAATTTCTCGACCTGTGGAGGAGGTACAGTTTTCACACCGT 1653
QY 1624 CAGTGACCAACCCACCTGCCCAAGCGGGCAAGTACTGCGAGGAGTGAGGTGAAA 1683
DB 1654 GAGTGCAGAGACCCGAGCTCAGATGGAGGAAGATACTGCTGGGTGGAGAGCCAG 1713
QY 1684 TACCGATCTCAACCTGGAGCCCTGCCAGCTCAGCCTCCGGAAGAGCTTCCGGGAG 1743
DB 1714 TACCAGTATCCACAGCGAGGATGCC-----CCTGACGGGAAAGCTTCAGGAG 1767
QY 1744 GAGCAGTGTGAGGCTTTCAGAGGCTTACACAGCAGACCAAGCGCTCACTCTCGCGGTG 1803
DB 1768 CAGCAGTGTGAGAGATATATGCTTACAAATTACACTGACATGGACGGGAATCT---CCTG 1824

QY 1804 GCATGGTCCCAAGTACTCCGGCGTGTCTCCCGGACAAAGTCAAGCTCATCTGCGCA 1863
DB 1825 CAGTGGTCCCCAAGTATGTGGGTGTCCCGCGGACCGCTGCAAGTGTCTTTCGCGCA 1884
QY 1864 GCAATGGCACTGGTACTTCTATGTGTGGCACCAAGGTGTGAGCGGACGCTGTGC 1923
DB 1885 GCGGGGAGAGGAGGTCAAAGTGTTCAGGCGCAAGGTGATGTGACACCTGTGT 1944
QY 1924 TCTCTGACTCCACCTCCGTCTGTGTCCAAAGCAAGTGCATCAAGGTGTGTATGGG 1983
DB 1945 GGGCAGAAACACTGCGCATCTGTGTCCGTGGCAGTGTCAAGCGCGCTGTGACCAT 2004
QY 1984 AACTTGGGCTCCAAAGAGATTCACAAGTGTGGGTGTGTGGGGGAGACATAGAGC 2043
DB 2005 GTGTGGACTCCGCTCGGAAGCTGGACAAATCGGGGTGTGTGGGGGCAAGGCAACTCC 2064
QY 2044 TCAAGAAGGTGACTGGGACTCTTCAAGCCCATGCTGCTCAAAATTTCTGTGTGGCC 2103
DB 2065 TCGAAGAGGTCTCCGGGTCCCTCACCCCAACAAATATGGCTACATGACATGTGCACC 2124
QY 2104 ATCCCGCAGGCGCTCAAGCATCGACATCCCGCAGCGGTTTACAAAGGCTGTATCGGG 2163
DB 2125 ATCCAGCTGTGTCACATAATATGACGTGAAGCAGCAGCGGACCCGGGTGTGCAGAAC 2184
QY 2164 GATGACAACTACTGGCTCTGAAGACAGCAGCAGCAAGTACCTGCTCAACGGGCAATTTC 2223
DB 2185 GATGGAATACTCTCGGCTGAAGAGCGGTGATGGGAGTACCTGCTCAACGGCAACTTC 2244
QY 2224 GTGTGTCCGGGTGGAGGGGACCTGTGTGTGAAGGCGAGTCTGCTGGGTACAGCGGC 2283
DB 2245 GCAATCTCTGCCATAGAGCAGGACATCTGTGTGAAGGGGACCATCTGAGTACAGCGGC 2304
QY 2284 ACGGCGACAGCGGTGGAGAGCTGCAGGCTTCCCGGCCCATCTCTGGAGCGGTGACCGTG 2343
DB 2305 TCCATCGCCACCTCGAGCGCTGCAGAGCTTTCGGCGCTTCCAGAGAGCTCTGACAGTG 2364
QY 2344 GAGGTCTCTCCGT---GGGGAAGATGACACCGCCCGGTCGCTACTCTCTTATCTG 2400
DB 2365 CAGTCTCTGACAGTCCCTGGGAGGTCTTCCCGGCAAGTACAAATACACTCTTCTTGT 2424
QY 2401 CCCAAGA 2408
DB 2425 CCTAATGA 2432
RESULT 12
E58655
LOCUS
DEFINITION Novel metallic protease.
ACCESSION E58655
VERSION E58655.1 GI:18629877
KEYWORDS JP 2001017183-A/3.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2670)
AUTHORS Yamaji, N., Nishimura, K. and Sasamata, M.
TITLE Novel metallic protease
JOURNAL Patent: JP 2001017183-A 3 23-JAN-2001;
YAMANOUCHI PHARMACEUT CO LTD
COMMENT OS Homo sapiens (human)
PN JP 2001017183-A/3
PD 23-JAN-2001
PF 09-JUL-1999 JP 1999196584
PR NOBORU YAMAJI, KOICHI NISHIMURA, MIHO SASAMATA
PC C12N15/09, C07K16/40, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC
C12N9/50, C12O1/37,
PC C12N15/00, C12N5/00
CC
FH Key Location/Qualifiers
FT source 1. .2670

FEATURES	FT	Location/Qualifiers	/organism="Homo sapiens (human)"
source		1..2670	
BASE COUNT	511 a	853 c	867 g
ORIGIN		439 t	
Query Match	23.3%	Score 666;	DB 6; Length 2670;
Best Local Similarity	59.2%	Pred. No. 6.6e-109;	
Matches 1355;	Conservative	0; Mismatches 885;	Indels 48; Gaps 11;
QY	157	CTCATTTTTCAGATCACAGCATTTTCAGAGACATTTTACCTACACCTGACCCGGATCGT	216
Db	157	CTCGCGCTCCACCTGTCGCGCTTCGGCAAGGGCTTCGTGCTGGCGCTGGCGCCGACGAC	216
QY	217	CAGTTCTTGGCTTCGCGCTTCTCCACTGAGCATCTGGCGCTCCCGCTCCAGGGGTACCC	276
Db	217	AGCTTCTTGGCGCGCGAGTTCAAGATCAGCGCTTCGGGGCTCCGGCGCGGCG---ACC	273
QY	277	GGGGCTCTTCAGAGCTCGAGAGCTGCTTCTATCTCGGGAGCTGAACCGCGAGCCGGAC	336
Db	274	GGGGCGAGCGGGGCTCGCGGCTGCTTCCTCCGGCACCGTAATGGGGAGCCCGAG	333
QY	337	TCGTTTCGCTGTGTGAGCGCTGTGCGGGGGCTTCGCGAGCGTTTGGCTACCGAGGCGCC	396
Db	334	TCGCTGGCGGGTTCAGCGCTGTGCGCGGGCTGAGCGGCTCTCTCTCTCTGTCGACGGCAG	393
QY	397	GAGTATGTCATTAGCCCGCTCCCACTAGTCAGCGCGCGGGCGCGACCGCAACAGCCAG	456
Db	394	GAGTTCACCATCCAGCGCGAGGGCGGGGGCTTCCTGGCTCAGCGCACCGCGCTGCAG	453
QY	457	GGCGCACACCTTCTCCAGCGCGGGGGTTCGCGGGCGGGCTTCGCGAGACCCACCTCT	516
Db	454	CGCTGGGTCCCGCGGAGCGCCGCCCTCCCGGAGGACCCGAGTGGAGGTGGAGCG	513
QY	517	CGCTGGGGTG- GCCTCGGCTGGAACCCCGCATCTCTAGGGCGCTTGGACCTTACAA	575
Db	514	GGAGAGGCTCAGAGCGAGGAGAGAGACACACAGGAGGACGACGAGGAGAGAGCCAA	573
QY	576	GCCGCGCGGGCGGGCTTCGGGGAGAGTCGTGAGCGGCGCAGG-----TCGTGGCGCGCC	630
Db	574	GAAGAGGCGCAGAAGCGCTAGCGAGCGCCACCGCCCTGGGGGCCACAGTAGGACC	633
QY	631	AAGCGTTTCGTGCTATCCCGCGTACGTGAGACGCTGTGTGTCGGCGAGAGTCAATG	690
Db	634	AAGCGGTTGTGTCTGAGCGCGCTTCGTGAAGACGCTGCTGTGTGGCGCATGTCTCATG	693
QY	691	GTCAAGTTCACGCGCGGACCTCGAATATTCGTGAGCGTCTGGCAACGCGCGCG	750
Db	694	GCTGCTTCTACGGGGCGGACCTCGAACCACATCTGAGTTAATGTCTGTGGCAGCC	753
QY	751	CGACTCTACGCCCATCCAGATCCTCAACCCCATCAACATCTGTGTGGTCAAGGTCGTG	810
Db	754	CGAATCTACAGCACCCGACGATCAAGAATTCATCAACTGATGGTGGTAAAGTCTG	813
QY	811	CTTCTTAGAGATGTGACTCCGGGCCCAAGTCAACGGAATGAGTGGGGCCCTGACGCTCGC	870
Db	814	ATCGTAGAGATGAAAATGGGGCCAGAGGTGTCCACAATGGGGGGCTTTACACTCGCT	873
QY	871	AACCTTGTGCTGGCAGAGAAGCTGAACAAAGTGAAGTGAAGACACCCCGAGTACTGG	930
Db	874	AACCTTCAACTGGCAGCGCGTTCACACGCGCCAGCGACCCGACACGACCATAC	933
QY	931	GACACTGCCATCTCTTACACGAGCAGCACTGTGTGG-----AGCCACCACTGTGACACC	987
Db	934	GACACGCCCATCTGCTTCACGACAGCAACTTCTGTGGCAGGAGGGGCTGTGTGACACC	993
QY	988	CTGGCATGCTGATGTGGGTACCATGTGTGACCCCAAGAGAGCTGCTCTGTCTATGAG	1047
Db	994	CTGGGTGCGACACATCGGGACCATTTGTGACCCCAAAAAGCTGCTCCGTGATCGAG	1053
QY	1048	GACGATGGGCTTCCATAGCCTTTCACACTTCGCCACGAGCTGGGCCACGTTTCAACATG	1107

Db 2125 ATCCAGCTGGTCCCACTAATATGACGTGAAGCAGCGGACCCACCGGGTGTGCAGAAC 2184

Qy 2164 GATGACAACTACCTGGCTCTGAAGAACAGCAGCAAGGCAAGTACCTGCTCAACGGGCAATTC 2223

Db 2185 GATGGGAACCTACCTGGCTCTGAAGAACAGCGGTGATGGCGAGTACCTGCTCAACGGCAACCTG 2244

Qy 2224 GTGCTGTGCGGCTGGAGCGGACCTGGTGGTGAAGGGCAGTCTGCTGGCGGTACAGCGGC 2283

Db 2245 GCCATCTCTGCCATAGCAGGACATCTTGGTGAAGGGGACCATCTCTGAAGTACAGCGGC 2304

Qy 2284 ACGGGACACGGTGGAGGACCTGGAGGCTTCCCGGCCCATCTCTGGAGCGCTGACCGTG 2343

Db 2305 TCCATGCCACCTGGAGCGCTGAGAGCTTCCGGCCCTTCCAGAGCCTCTGACAGTG 2364

Qy 2344 GAGTCTCTCCGT---GGGGAAGATGACACACCGCCCGGGTCCGCTACTCTCTTCTATCTG 2400

Db 2365 CAGTCTCTGACAGTCCCTGGCGAGGCTTCCCGCCCAAGTCAAAATACACCTCTTGT 2424

Qy 2401 CCCAAAGA 2408

Db 2425 CCTAATGA 2432

RESULT 13

AF060153 3711 bp mRNA linear PRI 11-AUG-1999

LOCUS Homo sapiens METH2 protein (METH2) mRNA, complete cds.

DEFINITION AF060153

ACCESSION AF060153.1 GI:5725507

VERSION

KEYWORDS

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 3711)

AUTHORS Vazquez,F., Hastings,G., Ortega,M.A., Lane,T.F., Oikemus,S., Lombardo,M., and Iruela-Arispe,M.L.

TITLE METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity

JOURNAL J. Biol. Chem. 274 (33), 23349-23357 (1999)

MEDLINE 99367466

PUBMED 10438512

REFERENCE 2 (bases 1 to 3711)

AUTHORS Vazquez,F., Hastings,G., Ortega,M.-A., Lane,T.F., Lombardo,M., Oikemus,S., and Iruela-Arispe,M.L.

TITLE Direct Submission

JOURNAL Submitted (16-APR-1998) Pathology, Beth Israel Deaconess Medical Center, 99, Brookline Avenue, Boston, MA 02214, USA

FEATURES

source 1..3711

/organism="Homo sapiens"

/db_xref="taxon:9606"

1..3711

/gene="METH2"

706..3378

/gene="METH2"

/note="extracellular protein with metalloprotease and TSP domains"

/codon_start=1

/product="METH2 protein"

/protein_id="A048081.1"

/db_xref="GI:5725508"

/translation="MFPAPAPRWLFLLLLLLLLPLARGAPAPAGGQASLVWP TRPVSAGELALHSAFGKGVFLRLAPDSFLAPEFKIERLGGSGRATGKGLKGF FSTVNGPELSEAVSLCRGLSGSLDGEFTIOQAGGSLAQPHRLQWRGPGAR PLRPGPEWETEGORQHEDSESEEEAEAGSEPPPLGATSRTKRFVSE ARVETLLVADASMAAFYQADLQNHILTLMSVAARIYKHPISKNLSLNVKVLIVD EKGPVSDNGGILTLEFNCNWRFPNPSDRHPEHYDTAILLTRONFCQEGCLDTLG VADIGTCIDPNKSCSVIEDEGLAAHTLAHELCHVLSMFDHDSKTRFLRPGWKHHV MAPFLVHLNLTLPWSPCSAMYLTELLDGGHGDCLLDAPGAALPLTGLRPAWLTOLD QOCRLQIFGDFRCPISADQVCAQWLTLDGAEPLCHTKNGSLWADPTGPGHLC SEGSLPEEVRPKVPVGDWAPWPGECSTRTGGVQFSHRECKDPDPPNGRYC LGRRAKYQSCTEECPDGKSFREQCEKYNNAYNTMDGNLLQWPKYAGVSPDRD

KLFCRARGSEFKVFEAKVIDTGLCGPETLAICVRGCVKAGCDHVVDSRKLKCGV CGKNGNSCRVSLQDTNYGNDIVTIIPAGATNIDVKORSHPGVNDGNYLAKTAD GQYLLNGLAISAEIDILVKYSGSIATLERLSRPLPEPLTVOLLVPGVEV FPKVKYTFVFNVDVDFSMOSSKERATTNIQPLHLAQWVLGWSWSECSTCGAGWRR TVECRDPGSOASATCNKALPKAPCESQLCPL"

BASE COUNT 733 a 1126 c 1201 g 646 t 5 others

ORIGIN

Query Match 23.3%; Score 666; DB 9; Length 3711;

Best Local Similarity 59.2%; Pred. No. 6.2e-109;

Matches 1355; Conservative 0; Mismatches 885; Indels 48; Gaps 11;

Qy 157 CTCATTTTTCAGATCACAGATTTTCAGAGGACATTTTACCTACACCTGACCCCGATGCT 216

Db 865 CTCGGCTCCACCTCTCCGCTTCGGCAAGGGCTTCGTGTTGCGCGCTGGCCCGAGAC 924

Qy 217 CAGTTCTTGGCTCCGCTTCCTCACTGAGCATCTGGCGCTCCGCTCCAGGGGCTCAC 276

Db 925 AGCTTCCTGGCGCCCGAGTTCAAGATCAGCGCTCGGGGCTCCGGCCGGCG---ACC 981

Qy 277 GGGGCTCTTCAGACCTCGACGCTGCTTCTATTCTTGGGGACGTGAACGCGCGAGCGGAC 336

Db 982 GGGGCGAGCGGGGCTCGCGGCTGTTTTTCCCGCACCGTGAATGGGAGCGCGAG 1041

Qy 337 TCGTTCGCTGCTGAGCGCTGTGGGGGCTCCGCGAGGCTTTGGCTACCGAGCGCC 396

Db 1042 TCGCTGGCGGGCTGAGCGCTGTGGCGGGCTGAGCGGCTCTCTCTGCTGAGCGGAG 1101

Qy 397 GAGTATGTCATTAGCCCTGCTGCCAATGCTAGCGCGCGGGGCGGCGACGCAACAGCAG 456

Db 1102 GAGTTCACCATCCAGCGCGAGGGCGGGGCTCCCTGGCTCAGCGCGCACCGCTGCG 1161

Qy 457 GCGCACACCTTCTCAGCGCGGGGCTGTTCCGGCGGGGCTTCGGGAGACCCACCTCT 516

Db 1162 CGCTGGGCTCCGCGCGGAGCGCGCCCTCCGCGGAGGACCCGAGTGGGAGGTGGAGACG 1221

Qy 517 CGCTCGGGGTG-GCTCGGGCTGGAACCCCGCCATCTACGGGCGCTTGACCC---TACAA 575

Db 1222 GGAGAGGTTCAGAGCGAGGAGAGAGACCAACAGGAGGACGAGGAGAGAGAGACCAA 1281

Qy 576 GCGCGGGCGGGGCTTCGGGGAGAGTGTAGCGCGCGGCGAGG-----TC*GGGCGGCC 630

Db 1282 GAAGAGGAGGAGAGAGGCGCTAGCGAGCGCCACCGCCCTGGGGCGCACGAGTAGGACC 1341

Qy 631 AGCGTTTCGCTCTCTATCCCGCGGTACGTGGAGAGCGTGGTGGCGGAGAGTCAATG 690

Db 1342 AAGCGGTTGTCTGAGCGCGCTTCGTGGAGAGCGTGGTGGCGCGATGCTGCTCAATG 1401

Qy 691 GTCAAGTTCACGCGCGGACCTGGAACATTATCTGCTGACGCTGCTGGCAACGCGCGG 750

Db 1402 GCTGCTTCTAGCGGCGGACCTGCAGAACACATCTGACGTTAATGCTGTGCGACGCC 1461

Qy 751 CGACTCTACGGCCATCCCAAGCATCTCAACCCCATCAACATGTTGGTCAAGTGTG 810

Db 1462 CGAATCTACAAGCACCCAGCATCAAGAAATTCATCAACCTGATGTTGTTAAAGTGTG 1521

Qy 811 CTTCTTAGAGATCGTACTCCCGGCGGAGGTCACGGCAATGCGGCGCTGACGCTGCGC 870

Db 1522 ATCGTAGAAGATGAAATAATGGGGCGGCGAGGTTGTCGACAAATGGGGGCTTACACTGGT 1581

Qy 871 AACTTCTGTGCTCGCAGCAAGAGCTGAACAAAGTGAAGTGAACAGCACCCCGAGTACTGG 930

Db 1582 AACTTCTGCACTGGCAGCGGCGTTTCAACAGCGCCAGCGCCGCCACCCAGAGCACTAC 1641

Qy 931 GACACTGCCATCTCTTTCACAGGAGGACCTGTGTGG---AGCCACCACTGTGACACC 987

Db 1642 GACACCGCATCTCTGCTACACAGAGCAAACTTCTGTGGGCGAGGAGGGGCTGTGTACACC 1701

Qy 988 CTGGGCGATGCTGATGTGGGTACCATGTGTGACCCCAAGAGAAAGTGTCTCTCATTTGAG 1047

Db 1702 CTGGGCTGGCAGACATCGGGACCATTTGTGACCCCAACAAAGTGTCTCGTGATCGAG 1761

Qy 1048 GACGATGGGCTTCCATCAGCCTTACCCTGCTCCACAGAGCTGGGCCACAGTGTTCACATG 1107

Db	2833	ATCCAGCTGGTCCCACTAATATTGACGTGAAGCAGCGAGCCACCCGGGTGTCAGAAC	2892
Qy	2164	GATGACAACCTACCTGGCTCTGAAGAACACGCCAAGCAAGTACCTGCTCAACGGCGATTTC	2223
Db	2893	GATGGAAATCTACCTGGCGCTGAAGACGGCTGATGGCGAGTACCTGCTCAACGGCAACCTG	2952
Qy	2224	GTGGTGTGGCGGTGGAGCGGGACCTGGFHGTGAAGGCGAGTCTGCTCGGTACAGCGGC	2283
Db	2953	GCCATCTTGCCATAGACGACAGACATCTTGGTGAAGGGACCATCCTGAAGTACAGCGGC	3012
Qy	2284	ACGGGCACAGCGGTGGAGAGCTTGCAGGCTTCCCGGCCCATCTCTGGAGCGCCTGACCGTG	2343
Db	3013	TCATCGCCACCCCTGGAGGCGCTGCAGAGCTTCCGGCGCCTTGCACAGACCTCTGCACAGT	3072
Qy	2344	GAGTGCTCTCCGT---GGGGAAGATGACACCGCCCGGGTCCGCTACTCTCTTATCTG	2400
Db	3073	CAGCTCTGACAGTCCCTGGCGAGGTCTTCCCCCAAAAGTCAATACACCTTCCTTGTT	3132
Qy	2401	CCCAAGA	2408
Db	3133	CCTAATGA	3140

```

RESULT 14
D67076
LOCUS       D67076             4180 bp    mRNA    linear    F02 02-FEB-1999
DEFINITION  Mouse mRNA for secretory protein containing thrombospondin motifs,
              complete cds.
ACCESSION   D67076
VERSION     D67076.1  GI:1813339
KEYWORDS    secretory protein containing thrombospondin motifs; ADAMTS-1.
SOURCE      Mus musculus colon adenocarcinoma cell_line:murine colon 26 cDNA to
              mRNA.
ORGANISM    Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
              1 (sites)
AUTHORS     Kuno,K., Kanada,N., Nakashima,E., Fujiki,F., Ichimura,F. and
              Matsushima,K.
TITLE       Molecular cloning of a gene encoding a new type of
              metalloproteinase-disintegrin family protein with thrombospondin
              motifs as an inflammation associated gene
JOURNAL     J Biol Chem. 272 (1), 556-562 (1997)
MEDLINE     97150761
REFERENCE   2 (bases 1 to 4180)
AUTHORS     Kuno,K., Kanada,N., Nakashima,E., Fujiki,F., Ichimura,F. and
              Matsushima,K.
TITLE       Molecular cloning of a gene encoding a new type of
              metalloproteinase-disintegrin family protein with thrombospondin
              motifs as an inflammation associated gene
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 4180)
AUTHORS     Kuno,K.
TITLE       Direct Submission
JOURNAL     Submitted (28-SEP-1995) Kouji Kuno, Cancer Research Institute,
              Kanazawa University, Pharmaco-physiology, Takara-machi 13-1, Kanazawa,
              Ishikawa 920, Japan (Tel:0762-62-8151(ex.5454), Fax:0762-60-7704)
FEATURES   Location/Qualifiers
              1..4180
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /cell_line="murine colon 26"
              /cell_type="colon adenocarcinoma"
              1..4180
              /gene="ADAMTS-1"
              441..3296
              /gene="ADAMTS-1"
              /note="putative"
              /codon_start=1
              /product="secretory protein containing thrombospondin
              motifs"
              /protein_id="BAAL1088.1"
              /db_xref="GI:1813340"
CDS
gene

```

/translation="MGDVQRAARSRLSNAHMLLLLLASITMMLLCARGAHRPTEED
 ELVLSLRAPGHDSTTLRLRDLAFQQLHLKLPDSGLFAPGFTLTQVGRSPCEA
 HLDPTDLAHCFYSGTVNGDPGSAALSLCEGVAFYLOGEEFFIOPAPGATSELA
 PAVPEESARQPHILRRRRRGSGAKGVMDDETLPSTDSRPESONTQWVPRD
 TPQDAGSGSPGIRKRFVSPRYVETMLVADSMADFHGSLKHLTLTLFVAAAR
 YKHSIRNSISLVVVKILVIEEKGPEVTSNAALTLLNFCSWQKOHNSPDRPEHY
 DTAIIFTQDLGSHGSDTLMDADVGTCDPSRSCVIEDDGLQAFTTAHELGHVN
 MPHDDAKHSDVSGDLSHMLSLSDHSQWSPCSAYMVTSLDNGHGEIMDK
 PNPILKPLSDPLGLYDANROQETTFGEESKCPDAASTCTLMCTGTSGLLVCOTK
 HPWADGTSCGEGKVCWCKNKTMDKHFAFVHSGWPMGWDGSRCTGGGVOT
 MRECDNPVKGKGYCEKRVRYRSCNIEDCPDNGKTFREOCEANEFKSAFGNE
 PIVEWTPKYAGVSPDKCKLCEAKGICYFVLPKPYVDGTPCPSDSTSVQCVQK
 AGCDRIIDSKKFKDCKGCGGNGSTCKKMSGIVSTRPYDHIIVIPAGATNIEVKR
 NORGSRNGSLFIAHADGTYLILNGFNFTLSLEODLYKTVLRYSGSSAALERISF
 SPKLEPLTQVLMVGHALRPKIFYFMKKTFESNAIPFSEWVIEEWGSCSTCS
 GQWRVVOQDINGHASECAKEVPKASTPCADLPCHPMQVGDWSPCSKTCGRGYK
 RFLKCVSHDGGVLSNESCDPLKPKHYIDFCTLTQCS"
 BASE COUNT 1078 a 1051 c 1146 g 905 t
 ORIGIN

Query Match 22.3%; Score 636; DB 10; Length 4180;
 Best Local Similarity 60.7%; Pred. No. 1.3e-103;
 Matches 1098; Conservative 0; Mismatches 700; Indels 12; Gaps 3;

QY 631 AAGCGTTTCGTCTATCCGCGGTACGTGGAGAGCGTGGTGGTGGCGGACAGTCAATG 690
 DB 1143 AAGCGATTTGTCTAGAGCCCCCGTTATGTGGAAACCATGCTCGTAGCTGACACGACCGATG 1202
 QY 691 GTCAAGTTCCAGCGCGGACCTGGAACATTATCTGTGACGCTGTGTCGACGCGGCGG 750
 DB 1203 GCGACTTCCAGCGGACGGTCTAAGACATTACCTCTAACCCTGTCTCGTGGCAGCC 1262
 QY 751 CGACTTACCGGATCCCAAGCATCTCAACCCCATCAACATCGTTGTGTGTCAGGTGCTG 810
 DB 1263 AGGTTTTTACAAGCATCCCAAGCATTAGGAATTCAATTAGCTGTGGTGGTGAAGATCTG 1322
 QY 811 CTCTTTAGAGATCGTACTCCGCGCCCAAGTCACCGCAATGCGGCGCTGACGCTGGC 870
 DB 1323 GTCATATAGAGGAGGACGAGGACGAGAGTTACTCTCAATGACGCTCTCACCCCTCGG 1382
 QY 871 AACTTCTGTGCTGGCAGAGAGCTGAACAAAGTGAAGTGAACAGCAACCCCGAGTACTGG 930
 DB 1383 AATTTCTGACGTGGCAGAAACACACACACCCAGTGACCGGGATCCAGAGCACAT 1442
 QY 931 GACACTGGCATCTCTTACAGGAGGACCTGTGTGGAGCCACACCTGTGACACCCGTG 990
 DB 1443 GACACTGCAATCTGTCTACAGAGAGGATTTATGTGGCTCCACACGCTGTGACACTCTC 1502
 QY 991 GGCATGGCTGATGTGGGTACCATGTGTACCCCAAGAGAGCTGCTGTCTATTGAGGAC 1050
 DB 1503 GGNATGGCAGATGTTGGAAACCCGTATGTGACCCACAGAGGCTGCTCAGTCATAGAAT 1562
 QY 1051 GATGGCTTTCCATGACGCTTACCACTGCGCCACGAGCTGGGCGACGCTGTTCAACATGCC 1110
 DB 1563 GATGGTTTCAAGCGGCTTACCAAGCCCATGAATTTGGGCCATGTGTTAATCATGCCG 1622
 QY 1111 CATGACATGTGAAGTCTGTGAGGAGGTGTTTGGGAAGCTCCGAGCCCAACACATGATG 1170
 DB 1623 CACGATGATGCTAAGCACTGTGCCAGCTTGAATGGTGTGAGTGGCGATTCATCTGATG 1682
 QY 1171 TCCCGGACCTCATCAGATCAGCCTGCAACCCCTGTGACGCTGAGTGTGCTGCCATC 1230
 DB 1683 GCTCGATGCTTCCAGCTTACGACCATAGCCAGCCCTGGTCACTTGGAGTGCCTACATG 1742
 QY 1231 ATCAACGACTTCTGGACAGCGGCGACGTGACTGCCCTGCTGGACCAACCCAGCAAGGCC 1290
 DB 1743 GTCAGCTCTCTAGATAATGACACGGGGAATGTTGATGGACACAGCCCGCAGATCCA 1802
 QY 1291 ATCTCCCTGCCAGATGTCGCGGCGCGCAGCTACACCCCTGAGCCAGCAGTGGAGGTG 1350
 DB 1803 ATCAAGCTCCCTCTGATGCTTCCCGGTACCTTGTACGATGCAACCGCCAGTGTCAAGTT 1862
 QY 1351 GCTTTTGGGCTGCCAAGCCCTGTCTTACATGCAAGTAC---TGACCAAGCTGTGG 1407

DB 1863 ACATTCGGAGAGGAATCCAAAGCACTGCCTCTGATCGACCCAGCACATGTACTACCTGTGG 1922
 QY 1408 TGCACCGGAAGGCCAAGGGACAGATGTGTGTGGCCAGACCCGCCACTTCCCTGSGGCCGAT 1467
 DB 1923 TGCACCTGCACCTCGGTGGCTTACTGTGTGGCCAAACAAACACTTCCCTTGGGACAGAT 1982
 QY 1468 GGCACACGCTGTGGCGAGGCAAGCTCTGCCTCAAGGGGCGCTCGTGGAGAGACACAAC 1527
 DB 1983 GGCACACGCTGTGGAGAGGGAAGTGTGTGTAGTGGCAAGTGCCTGTA/CAAGACAGAC 2042
 QY 1528 CTCACACAGCACAG-----GGTGGATGGTTCCTTGGGCCAAAATGGGATCCCTATFGCCCC 1581
 DB 2043 ATGAAGCATTTTGTACTCTCTCTGTTTATGGAAGCTGGGGACCATGGGACGCTGGGAGAC 2102
 QY 1582 TGCTTCGGCACATGTGTGGGGGCGTGCAGTGCAGGAGGAGGACGTCACCAACCCCAACC 1641
 DB 2103 TGCTCAAGAACCTGTGTGGTGGAGTTCAATACATATGAGATGTGACACCCCAAGTC 2162
 QY 1642 CCTGCCAACGGGGCAAGTACTTCGAGGGAGTGGAGTGAATACCGATCCTGCAACCTG 1701
 DB 2163 CCAAGAACGGAGGAAGTACTGTGAAGGCAACAGAGTCCGCTACAGGTCTCTGTAACATC 2222
 QY 1702 GAGCCCTGCCCACTCAGCTCCGGAAGAGCTTCCGGGAGGAGCAGTGTGAGGCTTTC 1761
 DB 2223 GAGGACTGTCCAGACAATAA---CGGAAACGTTTCAGAGAGGAGCAGTGCAGGGCGCAC 2279
 QY 1762 AACGGCTACACACACAGCAGCCGCTACTCTCGCGTGGCATGGTGGTCCCAAGTAC 1821
 DB 2280 AATGATTTTTCCAAAGCTTCTTTTGGGAATGAGCCACTGTAGAGTGGACACCCAGTAC 2339
 QY 1822 TCCGGGTGTCTCCCGGGGACAGTGAAGTCAATTCGCGAGCAATGACATGCGGTAC 1881
 DB 2340 GCGGCGCTCTCGCCAAAGAGCAGGTGCAAGCTACCTGTGAAAGCAAAAGCATTTGGCTAC 2399
 QY 1882 TTCTATGTGTGGCCACCAAGGTGTGAGGAGGACGCTGTCTCTGTGACTCCACTCC 1941
 DB 2400 TTTTGTGCTTACAGCCCAAGTGTGTGAAAGCTGGCTGTGATCGCATCATAGACTCTACCTCT 2459
 QY 1942 GTCTGTGTCCAAGCAAGTGCATCAAGGCTGGCTGTGTGATGGGAACCTTGGGCTCCAAAG 2001
 DB 2460 GTCTGTGTGCAAGGCAAGTGTGTGAAAGCTGGCTGTGATCGCATCATAGACTCTCCAAAG 2519
 QY 2002 AGATTCACACAGTGTGGGTGTGTGGGGGAGACATTAAGAGCTGCAAGAGGTGACTGGA 2061
 DB 2520 AAGTTTGAATAGTGTGGCTTGTGGAGAAACGGTTCCACATGCAAGAAAGATGTCAAG 2579
 QY 2062 CTCTTTCACCAAGCCATGCTATGCTACAAATTTGCTGTGGCCATCCCGCAGCGCTCA 2121
 DB 2580 ATAGTCACTAGTACAAAGCTGGGTATCATGACATTTGTCACATTTCTGTGGAGCCACC 2639
 QY 2122 AGCATTCACATCCCGCCAGCGGTTACAAAGGCTGATCGGGGATGACAAATACCTGGCT 2181
 DB 2640 AACATTAAGTGAACATCGAATCAAGGGGTCCAGAAACAATGACAGCTTCTTGGCT 2699
 QY 2182 CTGAAGAACCCCAAGCAAGTACTCTCTCAAGGGGCAATTTGCTGTGGGTGGGAGGAG 2241
 DB 2700 ATTAGAGCCGCTGATGTGTACTATTTCTGAATGGAACCTTCACTCTGTCCACACTAGAG 2759
 QY 2242 CGGACCTGTGTGTGAGGAGGAGTCTGCTGGGTACAGCGGACGCGGACAGCGGAGGAG 2301
 DB 2760 CAAGACCTTACCTTACAAGGTACTGTCTTAAGGTACAGTGTCTTCCGGCTGGGAG 2819
 QY 2302 AGCCTGACAGGCTTCCCGGCGCCATCTTGGAGCCGCTGACCGTGGAGGTCTCTCCGCTGGG 2361
 DB 2820 AGATCCGACAGCTTATGTCACCAAGAACCCCTTAACCATCCAGGTCTTATGCTAGGC 2879
 QY 2362 AAGATGACAGCGGCGGCTCCGCTACTCTTCTATCTGCCCAAGAGCCTCGGGAGGAC 2421
 DB 2880 CATGCTCTCGGACCAAAATTAATTCACCTACTTTATGAAAGAAAGACAGAGTCAATTC 2939
 QY 2422 AAGTCTCTCTC 2431

Db 2940 AACCCATTC 2949

RESULT 15
AF207664 4659 bp mRNA linear PRI 10-JAN-2000
LOCUS Homo sapiens matrix metalloprotease (ADAMTS1) mRNA, complete cds.
DEFINITION AF207664
ACCESSION AF207664
VERSION AF207664.1 GI:6685071
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4659)
AUTHORS Glienke,J., Schmitt,A., Pilarsky,C., Hinzmann,B., Weiss,B.,
Rosenthal,A. and Thierauch,K.H.
TITLE Genes differentially expressed by endothelial cells in distinct
angiogenic states
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4659)
AUTHORS Glienke,J., Schmitt,A., Pilarsky,C., Hinzmann,B., Weiss,B.,
Rosenthal,A. and Thierauch,K.H.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-1999) Experimental Oncology, Schering AG,
Muellerstr 178, Berlin 13342, Germany
FEATURES
Location/Qualifiers
1..4659
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="endothelial cells"
1..4659
/gene="ADAMTS1"
1..442
/gene="ADAMTS1"
443..3346
/gene="ADAMTS1"
/note="METH-1-like protein"
/product="matrix metalloprotease"
/protein_id="AAF2372.1"
/db_xref="GI:6685072"
/translation="MQRAVPEFGRRKLGSDMGNAERAPSRSPFVPTLLLAALL
AVSDALGRESEDELVPLEPAPCHGTRRLRLHAFDQDLRLPDSSEFLAPFTLL
QNVGRKSGSETPLPDLAHCFYSGTVNGDPSAAALSLCEGVGAFVLYLGEAYIQP
LPAASRLATAAPGEKPPAPLOFPHLLRNRODVGTCGVVDDEPRPTGKAETDEDE
GTGEDEGAQSPQDPALOGVQPTGTSIRKRVSVSHRYVETMLVADQSAERHGS
GLKHLLTLFSAARLYKPSIRNSVLSVYVVKLVILVHDEQKFEVTSNAILFLRNFCN
WQKHPPSDRDAEHYDITAILFTRODLCGSQTCDLTMADVGTVCDDPSRSCSVIEDDG
LQAFTTAHELGHVFNPHDDAKQASLNGVNDMSHMASLSNLDSPQSPCSAYD
ITSFLDNHGECIMDKPNPIQLPDLPGTSYDANRQOFTFGEDSKHCPDAASTCST
LWGTGSGVLVLCQTKHFPWADGTCGEGKWCINGKCVNKTDRKHFDTPFHGSWMWG
PWGDCSRTCGGVQVYTMRECDNPVKNKGKCYCEKRVYRSCNLEDPCDNNKRTREE
QCEAHNEFSAFGSGPAVENIPKYAGVSPKDKCLICQAGIGFYFVLQPKVVDGTP
CPDSTSVQCGCVKAGCDRIIDSKKFKDCGVCVGGNGSTCKISGVSITAKPQYHD
IITPTGATNIEVKQRNQRNSGFLAKAAGDTYILNGDYTLSTLEQDITMYGVH
LRYSGSAALEIRSPSLKEPLTQVLTVGNALRPKIKYFVFKKKSEFNAPTFSS
AWYIEBWGSKSCSELGWRVLECRDINGOPASCEKVEKPASTRPCADHPQWQL
GEWSSCKTCKGKYKRSCLKSLSHDGGVLSHESDPLKKPKHFIDFCYMAECS"
3347..4659
/gene="ADAMTS1"
3'UTR
BASE COUNT 1242 a 1083 c 1255 g 1079 t
ORIGIN

Query Match 22.1%; Score 631.4; DB 9; Length 4659;
Best Local Similarity 57.1%; Pred. No. 8.7e-103;
Matches 1307; Conservative 0; Mismatches 911; Indels 69; Gaps 6;

Qy 175 GCATTTCCAGGAGACTTTTACCTACACCTGACGCCGGATGCTCAGTGTCTTGGCTCCCGC 234
|||||
Db 680 GCCTTTCACGAGCAGTGGATCTGGAGCTCGCGCCGACAGCAGCTTTTGGCGCCCGC 739
|||||
Qy 235 TTCTCCACTGAGCATCTGGCGGCTCCCGCTCCAGGGGCTCACCGGGGGCTC---TTCAGAC 291
|||||

Db 740 TTCACGCTCCAGACGTTGGGGCGCAAAATCCGGGTCCGAGACGCCGCTTCCGGAACCCGAC 799
|||||
Qy 292 CTGCGACGCTGCTTCTATTCTGGGAGACGTGAACGCCGAGCCGGACTCCTTCTGCTCTGTG 351
|||||
Db 800 CTGGGCGCATGCTTCTACTCCGCGACCGGTGAATGCGCATCCAGCTGGCTGCCCCCTC 859
|||||
Qy 352 AGCCTGTGGGGGGGCTCCGCGGAGCCTTTGGCTACCGAGGCGCGAGTATGTCATTAGC 411
|||||
Db 860 AGCCTCTGGAGGGGCTGGCGCGCCCTTCTACCTGCTGGGGGAGGCGTATTTCATCCAG 919
|||||
Qy 412 CGCTTGCCTCAATGCTAGCGCGC-CGGCGGCGCAGCGCAACAGCCAGGGCGCACACTTCT 470
|||||
Db 920 CGCTGCGCGCGCGCAGCGCGCTCGCCACCGCGCCCGCAGGAGAGCGCGCGCA 979
|||||
Qy 471 CCAGCGCGGGGTGTTCGGGGGGGCTTCCGGAGACCCACCTCTCTCGCTGGGGGGTGGC 530
|||||
Db 980 CCACCTACAGTTCCACCTCTCTCGCGCGGAATCGGAGGGCGAGCTCGGGGGCACGTGCGGG 1039
|||||
Qy 531 CTCGGGCTGGAAACCCCGCCATCTTACGGGCGCTGACACCTTACAAAGCGCGCGGGCGGG 590
|||||
Db 1040 GTCGTGGACGACGCGCGCGGCGGACTGGAAAGCGGAGACCGGAAGACGAGACGAAGG 1099
|||||
Qy 591 CTTTCGGGAGAGTCTAGCGCGCAGGTCTTGGGGCGC-----GCCAAGCGTTTGGTCTATCCGGGGTAC 627
|||||
Db 1100 ACTGAGGGCGAGGACGAAGGGGCTCAGTGTGCGCGCAGGACCCCGGCACTGCAAGGCGTA 1159
|||||
Qy 628 -----GCCAAGCGTTTGGTCTATCCGGGGTAC 657
|||||
Db 1160 GGACAGCCACAGGAACCTGGAAGACATAAGAAAGACGATTTGTCTCCAGTCAACGCTAT 1219
|||||
Qy 658 GTGGAGACGCTGGTGTGCGGACGAGTCAATGGTCAAGTTCCACGGCGCGGACCTGAA 717
|||||
Db 1220 GTGGAAACCATGCTTGTGGCAGCCAGTCCGATGGCAGAAATCCAGCGGAGTGGTCTAAG 1279
|||||
Qy 718 CATTATCTGCTGACGCTGTGGCAACGGCGCGGCGACTCTACCGGCATCCGAGCATCTC 777
|||||
Db 1280 CATTACCTTCTCACGTTGTTTTCGGTGGCAGCCAGATTGTACAAACACCCCGCAGCATTCGT 1339
|||||
Qy 778 AACCCCATCAACATGTTGTGTGTCAGGTGTGCTTCTTAGAGATCGTGACTCCGGGGCC 837
|||||
Db 1340 AATTGAGTTAGCTGGTGTGTGAAGATCTTGGTCTACCCAGCATGAACAGAGGGCGCG 1399
|||||
Qy 838 AAGGTACACCGCAATCGCGCCCTGACGCTGCGCAACTCTTGTGCTGGCAGAAAGCTG 897
|||||
Db 1400 GAAGTGACCTCCAATGCTGCTCTACTCTGCGGAATTTTGAACATGGCAGAGTAGCAC 1459
|||||
Qy 898 AACAAAGTGTGACAAAGCACCCCGAGTACTGGGACACTGCCATCCTCTTTCACAGGCG 957
|||||
Db 1460 AACCCACCCAGTGACCGGGATGCAGAGCACTATGACAGCAATTTCTTTTACCAGACAG 1519
|||||
Qy 958 GACCTGTGGAGCACCAACCTGTACACCCCTGGCGATGGCTGATGTGGTACCATGTGT 1017
|||||
Db 1520 GACTTGTGGGTCCACACATGTGATCTCTGGGATGGGTGATGTGGAACTGTGTG 1579
|||||
Qy 1018 GACCCCAAGAGAGCTGCTCTGCTATTGAGGACGATGGGCTTCCATCAGCTTCACTACT 1077
|||||
Db 1580 GATCCGAGCAGAGCTGCTCCGCTCATAGAAGATGATGTTTACAGCTGCTTTCACACCA 1639
|||||
Qy 1078 GCCCAGAGCTGGCGCAGTGTTCACATGCCCATGACAAATGTGAAAGTGTGTGAGGAG 1137
|||||
Db 1640 GCCCATGAATTAGGCCAGCTGTTTAACTGCCACATGATGATGCAAGCACTCTGCCAGC 1699
|||||
Qy 1138 GTGTTTGGGAAGCTCCGAGCCCAACCATGATGTCCTCCGACCCCTCATCCAGATGACCGT 1197
|||||
Db 1700 CTTAATGTGTGAACACGAGATTCACCATGATGGGTCAATGCTTTCACCTGAGCCAC 1759
|||||
Qy 1198 GCCAACCCCTGTGACGCTGACGTCTGCCATFACACCGACTTCTCTGACAGAGCGGCGAC 1257
|||||
Db 1760 AGCCAGCTTGTCTCTCTGAGTGCCTACATGATTAATCATTTCTGGATAATGGTAT 1819
|||||
Qy 1258 GTGACTGCTCTCTGGACCAACCCCGAGGCCATCTCCCTGCGCGAGAGTCTCCCGGGC 1317
|||||

```

Db 1820 GGGGATGTTTGTATGACAAAGCCTCAGAAATCCCATACAGTCCACAGGCGATCTCCCTGGC 1879
QY 1318 GCCAGCTACACCTGAGCCAGCAGTGGAGTGGCTTTTGGCTGGGCTCCAAAGCCCTGT 1377
Db 1880 ACTCGTACGATGCCAAGCGGAGTCCAGTTTACATTTGGGAGGAGCTCCAAACACTGC 1939
QY 1378 CCTTACATGACGTAC---TGCACCAAGCTGTGTGTGTCACCGGGAAGGCCAAGGACAGATG 1434
Db 1940 CCGGATGACGCCAGCACATGTAGCAGCTTGTGGTGTACCGGCACTCTGTGTGGGTGCTG 1999
QY 1435 GTGTGCCAGACCGCCGCTCCCTGGCGATGGGACCAAGCTGTGGGAGGCGCAAGCTC 1494
Db 2000 GTGTGTCAAACCAACACTTCCCGTGGGCGATGGGACCAAGCTGTGGGAGGGAATGG 2059
QY 1495 TGCTCTAAAGGGCCCTGCTGGAGACACAACTCTCAACAAG-----CACAGGTGGAT 1548
Db 2060 TGTATCAACGGCAAGTGTGTGAACAAACCGACAGAAAGCATTTTGATACGGCTTTTCAT 2119
QY 1549 GGTTCCTGGGCCAATGGGATCCCTATGGCCCTGCTGGGACACATGTGTGGGGCGTG 1608
Db 2120 GGAAGCTGGGGAAATGTGGGGCCCTGGGAGACTGTTCGAGAACGTGCGGTGGAGGAGTC 2179
QY 1609 CAGTGGCCAGGAGGAGTGCACCAACCCACCCCTGCCAAGCGGGGCAAGTACTGCGAG 1668
Db 2180 CAGTACAGATGAGGAATGTGACAAACCCAGTCCCAAGAAATGGAGGGAAGTACTGTGA 2239
QY 1669 GGAGTGAAGGTGAATACCGATCCTCAACCTGGAGCCCTGCCCAAGCTCAGCCTCCGA 1728
Db 2240 GGCAACGAGTGCCTACAGATCCTGTAACTTGAAGGACTGTCCAGAC---AATAATGA 2296
QY 1729 AAGAGCTTCGGGAGGAGAGTGTGAGGCTTTCACGGCTACACCAAGCAGCAACCGG 1788
Db 2297 AAAACCTTTAGAGAGGAACAATGTGAAGCACACACAGAGTTTTCAAAAGCTTCCCTTTGGG 2356
QY 1789 CTCACCTCGCCGTGGGTCGCCAAGTACTCCGGGTGTCTCCCGGGACAGAGTGC 1848
Db 2357 AGTGGGCTCGGTGGAATGGATCCCAAGTACGCTGGGCTCACCAGGACAGGTC 2416
QY 1849 AAGCTCATCTGCCGAGCCAAATGCACTGGCTACTTCTATGTGTGGCACCCCAAGTGTG 1908
Db 2417 AAGCTCATCTGCCAAGCCAAAGGCAATGGCTACTTCTCGTTTTGCAGCCCAAGTTGTA 2476
QY 1909 GACGGCAGCTGTGCTCTCTGACTCCACTCCGCTGTGTCTCAAGGCAAGTGCATCAAG 1968
Db 2477 GATGGTACTCCATGTAGCCAGATTCACCTCTGTCTGTGTGCAAGGACAGTGTGTAAA 2536
QY 1969 GCTGGCTGTGATGGGAACCTGGGCTCCAAGAGAGATTCGACAAAGTGTGGGTGTGTGG 2028
Db 2537 GCTGGTGTGATCGCATCATAGACTCCAAAGAAAGTTTGTAAATGTGTGTGGTGGGG 2596
QY 2029 GGAGACAATAAGAGCTGCAAGAGGTGACTGGACTCTTCAACCAAGCCCATGATGGCTAC 2088
Db 2597 GGAATGGATCTACTTGTAAAAAATATCAGGATCAGTTACTAGTCAAAACCTGGATAT 2656
QY 2089 AATTTGTTGGTGGCCATCCCGGAGGCGCTCAAGCATCGACATCCGCCAGCGCGGTAC 2148
Db 2657 CATGATATCATCAATTCCACTGGAGGCCCAACAGTGAAGTGAACAGCGGAACCCAG 2716
QY 2149 AAAGGGCTGATCGGGATGACAACTACCTGGCTCTGAAGAACAGCCCAAGCAAGTACCTG 2208
Db 2717 AGGGATCCAGGAACAATGGCAGCTTCTTGCCATCAAGCTGCTGATGGCACAATATT 2776
QY 2209 CTCACGGGCAATTCGTGTGTGTCGGGGTGGAGCGGACCTGGTGTGAAGGCGAGTCTG 2268
Db 2777 CTTAATGGTGACTACACTTTGTCCACCTTAGAGCAAGACATTTATGTACAAAGGTGTTGC 2836
QY 2269 CTGGGTACAGCGGCGAGCGGCTGGAGAGGCTGCAGGCTTCCCGGCCCATCCTG 2328
Db 2837 TTGAGGTACAGCGGCTCCTCTGCGGCAATGGAAAGAAATTCGACGCTTTAGCCCTCTCAA 2896
QY 2329 GAGCCGCTGACCTGGAGGCTCTCTCGGTGGGGAAGATGACACCGCCCGGCTCCGCTAC 2388
Db 2897 GAGCCCTTGACCATCAGGTTCTTACTGTGGCAATGCCCTTCGACCTAAATTAATAC 2956

```

QY 2389 TCCTTCT 2395

||| ||

Db 2957 ACCTACT 2963

Search completed: May 1, 2003, 22:35:26
Job time : 10450 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2003, 19:20:41 ; Search time 606 Seconds
(without alignments)
10602.230 Million cell updates/sec

Title: US-10-009-332-2

Perfect score: 2853

Sequence: 1 atgcttttgcgtggcgcctct.....gcgtcctgagccgtgctga 2853

Scoring table: IDENTITY_NUC

Gap 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-n-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2853	100.0	2853	22	AAH41003 Human metalloprotease
2	2846.6	99.8	2853	24	AAD35569 Human metalloprotease
3	2846.6	99.8	3446	24	AAD35571 Human metalloprotease
4	2814.4	98.6	2937	24	AAS97182 Human metalloprotease
5	2812.8	98.6	2930	24	ABK12894 Human metalloprotease
6	1216.6	42.6	1518	21	AAAS9827 Rat metalloprotease
7	1088	38.1	1104	24	AAD35570 Human metalloprotease
8	1045.2	36.6	1143	21	AAAS9831 Human metalloprotease
9	955.8	33.5	966	24	AAD35568 Human metalloprotease

10	666	23.3	2670	22	AAH20226	Human ADAM-type me
11	666	23.3	2670	22	AAF82166	Human ADAM type me
12	666	23.3	3008	20	AAZ32001	Human METH2 encodi
13	666	23.3	3008	22	AAC90058	Human METH2 coding
14	636	22.3	4180	20	AAZ32010	Human METH1 relate
15	636	22.3	4180	20	AAC90067	D67776 cDNA clone.
16	629.8	22.1	3261	22	AAZ32000	Human METH1 encodi
17	629.8	22.1	3261	22	AAC90057	Human METH1 coding
18	629.8	22.1	4014	19	AAV66508	Human integrin lig
19	628.8	22.0	2184	20	AAH17990	Human ADAMTS-1 cod
20	628.6	22.0	4712	20	AAH04374	Human secreted pro
21	628.2	22.0	2853	22	AAH20224	Human ADAM-type me
22	628.2	22.0	4676	20	AAH19955	Human fango-71 enc
23	618.8	21.7	4858	24	ABN86294	Human metalloprote
24	608.8	21.3	2346	21	AAAS9821	Human metalloprote
25	599.8	21.0	3126	24	ABN86300	Human metalloprote
26	592	20.8	3638	22	AAAF63441	Murine ADAMTS-8 CD
27	586.4	20.6	4192	20	AAH00725	Human aggrecanase
28	584.8	20.5	4301	22	AAH168146	Human metalloprote
29	584.8	20.5	4303	21	AAAS95826	Human metalloprote
30	583.2	20.4	4406	22	AAAS46100	Human DNA encoding
31	583.2	20.4	4407	21	AAA37111	Human PRO1563 (UNQ
32	583.2	20.4	4407	22	AAAF54426	Probe #46 used in
33	547.4	19.2	3250	20	AAH00726	Human aggrecan deg
34	545.8	19.1	3002	22	AAAF63437	Murine ADAMTS-5 CD
35	545.8	19.1	5530	21	AAAF75435	Human aggrecanase
36	530.2	18.6	1878	24	ABA98676	Nat aggrecanase co
37	523	18.3	2751	21	AAA95822	Rat metalloprote
38	471.4	16.5	2625	22	AAAF63444	Murine ADAMTS-9 CD
39	420.8	14.7	5053	24	ABK88916	Human aggrecanase
40	419.2	14.7	5605	24	ABK88915	Human aggrecanase
41	419.2	14.7	5808	24	AAAS97179	Human metalloprote
42	419.2	14.7	5949	24	AAAS98771	Human ADAMTS-SI nu
43	419.2	14.7	6033	22	AAD30577	Human metalloprote
44	417.6	14.6	4086	22	AAH49373	Human metalloprote
45	411.2	14.4	5804	22	AAAF63443	Murine ADAMTS-9 CD

ALIGNMENTS

RESULT 1	
AAH41003	ID AAH41003 standard; cDNA; 2853 BP.
XX	XX
XX	AAH41003;
XX	XX
DT	23-AUG-2001 (first entry)
XX	XX
DE	Human metalloprotease MDT56 cDNA.
XX	XX
KW	Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;
KW	osteopathic; antiarthritic; ss.
XX	XX
OS	Homo sapiens.
XX	XX
FT	Key Location/Qualifiers
FT	CDS 1..2853
FT	/*tag= a
FT	/product= "Metalloprotease"
FT	/note= "Aggrecanase activity"
XX	XX
PN	WO200134785-A1.
XX	XX
PD	17-MAY-2001.
XX	XX
PF	10-NOV-2000; 2000WO-JP07917.
XX	XX
PR	11-NOV-1999; 99JP-0321740.
PR	16-MAY-2000; 2000JP-0144020.
XX	XX
PA	(YAMA) YAMANOUCHI PHARM CO LTD.
PA	(KAZU-) KAZUSA DNA RES INST.


```
QY 1801 GTGCATGGGTGCCAAGTACTCCGGCGTGTCTCCCGGACAAAGTGCAGCTCATCTGC 1860
Db 1801 GTGCATGGGTGCCAAGTACTCCGGCGTGTCTCCCGGACAAAGTGCAGCTCATCTGC 1860
QY 1861 CGAGCCAATGGCACTGCTACTTCTATGTCTGCTGCACCCCAAGGTGTGCACGCGCTG 1920
Db 1861 CGAGCCAATGGCACTGCTACTTCTATGTCTGCTGCACCCCAAGGTGTGCACGCGCTG 1920
QY 1921 TGCTCTCTGACTCCACCTCCGCTGTGTCTCAAGGAAGTGCATCAAGCTGCCTGTGAT 1980
Db 1921 TGCTCTCTGACTCCACCTCCGCTGTGTCTCAAGGAAGTGCATCAAGCTGCCTGTGAT 1980
QY 1981 GGGAACTGGGCTCCAAAGAGAGATTCGACAAAGTGTGGGTGTGGGGGAGACATAAG 2040
Db 1981 GGGAACTGGGCTCCAAAGAGAGATTCGACAAAGTGTGGGTGTGGGGGAGACATAAG 2040
QY 2041 AGCTGCAAGAGTGTGACTGACTCTTTCACCAAGCCCATGCATGGCTACAAATTCGTGGTG 2100
Db 2041 AGCTGCAAGAGTGTGACTGACTCTTTCACCAAGCCCATGCATGGCTACAAATTCGTGGTG 2100
QY 2101 GCCATCCCCGAGCGCCTCAAGCATCGACATCCGCGAGCGGTTACAAAGGCTGATC 2160
Db 2101 GCCATCCCCGAGCGCCTCAAGCATCGACATCCGCGAGCGGTTACAAAGGCTGATC 2160
QY 2161 GGGATGACAACTACTTGGCTCTGAAGACAGCCAGGCAAGTACCTGTCAACGGGCAT 2220
Db 2161 GGGATGACAACTACTTGGCTCTGAAGACAGCCAGGCAAGTACCTGTCAACGGGCAT 2220
QY 2221 TTCTGTGTGTGGCGTGGAGCGGACCTGGTGTCAAGGGCAGTGTGCGGTACAGC 2280
Db 2221 TTCTGTGTGTGGCGTGGAGCGGACCTGGTGTCAAGGGCAGTGTGCGGTACAGC 2280
QY 2281 GGCACGGGCACAGCGTGGAGAGCCCTGCAGGCTTCCCGGCCCATCTGGAGCGCGTAC 2340
Db 2281 GGCACGGGCACAGCGTGGAGAGCCCTGCAGGCTTCCCGGCCCATCTGGAGCGCGTAC 2340
QY 2341 GTGGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTGGCTTCTCTTATCTG 2400
Db 2341 GTGGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTGGCTTCTCTTATCTG 2400
QY 2401 CCCAAGAGCTCGGAGGACAAAGTCTCTCATCCCAAGGACCCCGGGAGCCCTCTGTC 2460
Db 2401 CCCAAGAGCTCGGAGGACAAAGTCTCTCATCCCAAGGACCCCGGGAGCCCTCTGTC 2460
QY 2461 TTGCACAACAGCTCTCTCAGCTCTCCAAACAGGTGGAGCAGCGGACACAGGCCCCCT 2520
Db 2461 TTGCACAACAGCTCTCTCAGCTCTCCAAACAGGTGGAGCAGCGGACACAGGCCCCCT 2520
QY 2521 GCAGCTGGGTGCTGGCAGCTGGGGCGGTGCTCCGCGAGCTGGGCGAGTGGCCTGCAG 2580
Db 2521 GCAGCTGGGTGCTGGCAGCTGGGGCGGTGCTCCGCGAGCTGGGCGAGTGGCCTGCAG 2580
QY 2581 AAGCGGCGGTGACTGCCGGGCTCCGCGGAGCAGCAGCTCCCTGCCTGTATGCA 2640
Db 2581 AAGCGGCGGTGACTGCCGGGCTCCGCGGAGCAGCAGCTCCCTGCCTGTATGCA 2640
QY 2641 GCCCATCGGCCCTGGAGACACAAGCTCTGGGGAGCCCTGCCACCTGGGAGCTCAGC 2700
Db 2641 GCCCATCGGCCCTGGAGACACAAGCTCTGGGGAGCCCTGCCACCTGGGAGCTCAGC 2700
QY 2701 GCCTGTGCTACCTGTCTCAAGAGCTGGCGCGGGGATTTTCAGAGCGCTCCTCAAGTGT 2760
Db 2701 GCCTGTGCTACCTGTCTCAAGAGCTGGCGCGGGGATTTTCAGAGCGCTCCTCAAGTGT 2760
QY 2761 GTGGGCAAGGAGCGCGGTGTGGCCCGGAGCAGTGCACCTTGCACCGAAGCCCCAG 2820
Db 2761 GTGGGCAAGGAGCGCGGTGTGGCCCGGAGCAGTGCACCTTGCACCGAAGCCCCAG 2820
QY 2821 GAGCTGGACTTCTGCTCTCTGAGGCGGTGCTGA 2853
Db 2821 GAGCTGGACTTCTGCTCTCTGAGGCGGTGCTGA 2853
```

RESULT 2

```
AD35569
ID AAD35569 standard; cDNA; 2853 BP.
XX AC AAD35569;
XX DT 26-JUL-2002 (first entry)
XX DE Human protease cDNA #2.
XX KW Human; novel human protein; NHP; protease; biological disorder; obesity;
    high blood pressure; arthritis; connective tissue disorder; infertility;
    gene therapy; enzyme; gene; ss.
XX OS Homo sapiens.
XX FH Key
    CDS 1..2853
    FT /*tag= a
    FT /product= "Human protease #2"
    FT
    FT XX
    FT PN
    FT XX
    FT PD
    FT XX
    FT PF
    FT XX
    FT PR
    FT XX
    FT PA
    FT XX
    FT PI
    FT XX
    FT DR
    FT DR
    FT PT
    FT PT
    FT XX
    PS Claim 1; Page 35-36; 41pp; English.
    CC
    CC The present sequence is a cDNA encoding novel human protein (NHP),
    CC human protease. NHPs share structural similarity with animal proteases
    CC particularly zinc metalloproteases. Sequences of the invention are
    CC useful in therapeutic, diagnostic and pharmacogenomic applications.
    CC NHP polynucleotides are used as hybridisation probes for screening
    CC libraries and assessing gene expression patterns. They can also be
    CC used for treating related biological disorders such as obesity, high
    CC blood pressure, arthritis, connective tissue disorders and infertility.
    CC They are also used in gene therapy.
    XX
    SQ Sequence 2853 BP; 522 A; 920 C; 903 G; 508 T; 0 other;
```

```
Query Match 99.8%; Score 2846.6; DB 24; Length 2853;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2849; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 ATGCTTTTGTCTGGCATCCTTAACCTTCCCGGGGAAACCGCTGGAGCTTTGAG 60
Db 1 ATGCTTTTGTCTGGCATCCTTAACCTTCCCGGGGAAACCGCTGGAGCTTTGAG 60
QY 61 CCAGAGCGGAGGTAGTCTTCCCATCCGACTGGACCCGGACATTAACGGCTCCGCTAC 120
Db 61 CCAGAGCGGAGGTAGTCTTCCCATCCGACTGGACCCGGACATTAACGGCTCCGCTAC 120
QY 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCATTTTTCAGATCAGACATTT 180
Db 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCATTTTTCAGATCAGACATTT 180
QY 181 CAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCGCGCTTCC 240
Db 181 CAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCGCGCTTCC 240
```

241 ACTGAGCATCTGGGCGTCCCTCCAGGGCTCACCGGGGCTCTTCAGACCTGCGACGC 300
 241 ACTGAGCATCTGGGCGTCCCTCCAGGGCTCACCGGGGCTCTTCAGACCTGCGACGC 300
 301 TGCTTCTATTCTGGGACGTGAACGGCGGACCGGACTCGTTCGCTGCTGTGAGCCCTGTGC 360
 301 TGCTTCTATTCTGGGACGTGAACGGCGGACCGGACTCGTTCGCTGCTGTGAGCCCTGTGC 360
 361 GGGGGCTCCCGGAGCTTTGGCTACCGAGGCGCGAGTATGTCATTAAGCCGCTGCC 420
 361 GGGGGCTCCCGGAGCTTTGGCTACCGAGGCGCGAGTATGTCATTAAGCCGCTGCC 420
 421 AATGCTAGCGCGCGGGCGGCGAGCGGACAGCGGAGCGGACACCTTCTCCAGCGCGG 480
 421 AATGCTAGCGCGCGGGCGGCGAGCGGACAGCGGAGCGGACACCTTCTCCAGCGCGG 480
 481 GGTGTTCCGGCGGGCTTCCGGAGACCCACCTCTCGCTGCGGGGTGGCTCGGGCTGG 540
 481 GGTGTTCCGGCGGGCTTCCGGAGACCCACCTCTCGCTGCGGGGTGGCTCGGGCTGG 540
 541 AACCCCGCCATCTTACGGGCGCTGGACCTTACAAGCGCGGCGGGGCTTCGGGGAG 600
 541 AACCCCGCCATCTTACGGGCGCTGGACCTTACAAGCGCGGCGGGGCTTCGGGGAG 600
 601 AGTCGTAGCGCGGCGAGGTCTGGGCGGCGCAAGCTTTCGTATTCGCGCGGTACGTG 660
 601 AGTCGTAGCGCGGCGAGGTCTGGGCGGCGCAAGCTTTCGTATTCGCGCGGTACGTG 660
 661 GAGACGCTGTGTGCGGAGGAGTCAATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
 661 GAGACGCTGTGTGCGGAGGAGTCAATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
 721 TATCTGCTGAGCTGTGCGGAGCGGCGGCGGCTTACCGCATCCAGCATCTTCAAC 780
 721 TATCTGCTGAGCTGTGCGGAGCGGCGGCGGCTTACCGCATCCAGCATCTTCAAC 780
 781 CCCATCAACATCTGTGTGCTCAAGTGTCTTCTTAGAGATCGTACTCCGGGCGCAAG 840
 781 CCCATCAACATCTGTGTGCTCAAGTGTCTTCTTAGAGATCGTACTCCGGGCGCAAG 840
 841 GTCAACGCAATGCGGCGCTGACGCTGCGCACTTCTGCTGCTGCGGAGAGAGCTGAAC 900
 841 GTCAACGCAATGCGGCGCTGACGCTGCGCACTTCTGCTGCTGCGGAGAGAGCTGAAC 900
 901 AAATGAGTGACAAAGACCCCGAGTACTGGGACACTGCCATCTCTTACAGGCGGAGGAC 960
 901 AAATGAGTGACAAAGACCCCGAGTACTGGGACACTGCCATCTCTTACAGGCGGAGGAC 960
 961 CTGTGTGAGGCGCACCTGTGACACCTGCGCATGCTGCTGCTGCTGCTGCTGCTGCTG 1020
 961 CTGTGTGAGGCGCACCTGTGACACCTGCGCATGCTGCTGCTGCTGCTGCTGCTGCTG 1020
 1021 CCCAAGAGAGCTGCTGTGCTATTGAGGAGGATGGGCTCCATCAGCCTTCAACACTGCC 1080
 1021 CCCAAGAGAGCTGCTGTGCTATTGAGGAGGATGGGCTCCATCAGCCTTCAACACTGCC 1080
 1081 CACAGCTGGGCGAGCTTCAACATCCCATGCAATGTGAAGTCTGTGAGGAGGTG 1140
 1081 CACAGCTGGGCGAGCTTCAACATCCCATGCAATGTGAAGTCTGTGAGGAGGTG 1140
 1141 TTTGGGAGGCTCCGAGGCAACCATCATGCTCCCGACCTCATCCAGATCGACCGTGC 1200
 1141 TTTGGGAGGCTCCGAGGCAACCATCATGCTCCCGACCTCATCCAGATCGACCGTGC 1200
 1201 AACCCCTGGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
 1201 AACCCCTGGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
 1261 GACTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
 1261 GACTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
 1321 AGCTACACCTGAGCCAGGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380

1321 AGCTACACCTGAGCCAGCAGTGGCTTTTGGCGTGGGCTCCAAAGCCCTGTGCT 1380
 1381 TACATGCAAGTACTGACCAAGCTGTGTGACCGGAAAGGCAAGGGACAGATGGTGTGC 1440
 1381 TACATGCAAGTACTGACCAAGCTGTGTGACCGGAAAGGCAAGGGACAGATGGTGTGC 1440
 1441 CAGACCCGCCACTTCCCTGGGCGATGGCACAGCTGTGGCGAGGGCAAGCTCTGCCCTC 1500
 1441 CAGACCCGCCACTTCCCTGGGCGATGGCACAGCTGTGGCGAGGGCAAGCTCTGCCCTC 1500
 1501 AAAGGGGCTCGGTGGAGAGACAACTCAACAGCAGAGGGTGGATGGTTCCTGGGCC 1560
 1501 AAAGGGGCTCGGTGGAGAGACAACTCAACAGCAGAGGGTGGATGGTTCCTGGGCC 1560
 1561 AAATGGGATCCCTATGCGGCCCTGCTCGCGCACATGTGTGTGGGGCGTGCAGCTGGCCAGG 1620
 1561 AAATGGGATCCCTATGCGGCCCTGCTCGCGCACATGTGTGTGGGGCGTGCAGCTGGCCAGG 1620
 1621 AGGAGTGCACCAACCCACCCCTGCCAACGGGGCAAGTACTGCGAGGGAGTGAAGGTG 1680
 1621 AGGAGTGCACCAACCCACCCCTGCCAACGGGGCAAGTACTGCGAGGGAGTGAAGGTG 1680
 1681 AAATACCCATCTGCAACCTGAGGCCCTGCCCGAGCTCAGCTCCGGAACAGTTCCTCGG 1740
 1681 AAATACCCATCTGCAACCTGAGGCCCTGCCCGAGCTCAGCTCCGGAACAGTTCCTCGG 1740
 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTTCAACAGCAGCAGCAGCAGCAGCAGCAGCAG 1800
 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTTCAACAGCAGCAGCAGCAGCAGCAGCAGCAG 1800
 1801 GTGGCATGAGTGGGCGGCAAGTACTCCGGCGTGTCTCCCGGGACAGTCAAGCTCACTGCG 1860
 1801 GTGGCATGAGTGGGCGGCAAGTACTCCGGCGTGTCTCCCGGGACAGTCAAGCTCACTGCG 1860
 1861 CGAGCCATGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
 1861 CGAGCCATGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
 1921 TGCTCTCTGACTCCACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
 1921 TGCTCTCTGACTCCACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
 1981 GGGACCTGGGCTCCAAAGAGATTCGACAGTGTGGGCTGTGTGGGGGAGACAATAAG 2040
 1981 GGGACCTGGGCTCCAAAGAGATTCGACAGTGTGGGCTGTGTGGGGGAGACAATAAG 2040
 2041 AGCTGCAAGAAGGTGACTGGACTTTCACCAAGCCCATGCTGCTGCTGCTGCTGCTGCTGCTG 2100
 2041 AGCTGCAAGAAGGTGACTGGACTTTCACCAAGCCCATGCTGCTGCTGCTGCTGCTGCTGCTG 2100
 2101 GGCATCCCGCAGGGCGCTTCAAGCATCGACATCCGCGAGCGGCTTACAAAGGGCTGATC 2160
 2101 GGCATCCCGCAGGGCGCTTCAAGCATCGACATCCGCGAGCGGCTTACAAAGGGCTGATC 2160
 2161 GGGGATGACAACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
 2161 GGGGATGACAACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2220
 2221 TTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
 2221 TTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2280
 2281 GGCAGGGGACAGCGGTGGAGAGCCTGAGGCTTCCCGGCCCATCTCGAGCCCGCTGACC 2340
 2281 GGCAGGGGACAGCGGTGGAGAGCCTGAGGCTTCCCGGCCCATCTCGAGCCCGCTGACC 2340
 2341 GTGGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2400
 2341 GTGGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2400
 2401 CCCAAGAGCCTCGGAGGAGCAAGTCTCTATCCAAAGAGCCCGGGGAGCCCTCTGCTG 2460

Db 2401 CCCAAGAGGCTCGGAGGACAAAGTCTCTCATCCCAAGGACCCCGGGGACCTCTGTC 2460
QY 2461 TTGCACAAACAGCTCCCTCAGCTCTCCAAACAGGTGAGCAGCGGAGACAGGCCCCCT 2520
Db 2461 TTGCACAAACAGCTCCCTCAGCTCTCCAAACAGGTGAGCAGCGGAGACAGGCCCCCT 2520
QY 2521 GCAGCTGGGTGGCTGTCAGCTGGGGGCGTGTCCGCGAGCTGCGGAGTGGCCCTGCAG 2580
Db 2521 GCAGCTGGGTGGCTGTCAGCTGGGGGCGTGTCCGCGAGCTGCGGAGTGGCCCTGCAG 2580
QY 2581 AAGCGGCGGTGGACTCCCGGGGCTCCGCGGGGAGCGGACGCGACGTCCTGCTGTATGCA 2640
Db 2581 AAGCGGCGGTGGACTGTCGGGGCTCCGCGGGGAGCGGACGCGACGTCCTGCTGTATGCA 2640
QY 2641 GCCCATCGGCGCTGGAGACACAAAGCTGCGGGGAGCCCTGCCACCTGGGAGCTCAGC 2700
Db 2641 GCCCATCGGCGCTGGAGACACAAAGCTGCGGGGAGCCCTGCCACCTGGGAGCTCAGC 2700
QY 2701 GCCTGGTACCTGCTCCAAAGAGTGGCGCGGGGATTTTCAGAGGCGCTCACTCAAGTGT 2760
Db 2701 GCCTGGTACCTGCTCCAAAGAGTGGCGCGGGGATTTTCAGAGGCGCTCACTCAAGTGT 2760
QY 2761 GTGGGCCACGAGGCGCGCTGCTGGCCCGGGACCAAGTGAACCTTGACCGCAAGCCCGAG 2820
Db 2761 GTGGGCCACGAGGCGCGCTGCTGGCCCGGGACCAAGTGAACCTTGACCGCAAGCCCGAG 2820
QY 2821 GAGTGGACTTCTCGCTCCCTGAGGCGGTGCTGA 2853
Db 2821 GAGTGGACTTCTCGCTCCCTGAGGCGGTGCTGA 2853

RESULT 3

AAD35571

ID AAD35571 standard; cDNA; 3446 BP.

AC AAD35571;

XX AAD35571;

XX 26-JUL-2002 (first entry)

XX Human protease cDNA #4.

XX Human; novel human protein; NHP; protease; biological disorder; obesity;

KW high blood pressure; arthritis; connective tissue disorder; infertility;

KW gene therapy; enzyme; ss.

XX Homo sapiens.

OS W0200226949-A2.

PN 04-APR-2002.

XX 27-SEP-2001; 2001WO-US30350.

XX 29-SEP-2000; 2000US-236689P.

XX (LEXI-) LEXICON GENETICS INC.

XX Friddle CJ, Hilbun E;

XX WPI; 2002-372123/40.

XX Novel nucleic acid encoding a human protease, useful as a hybridization

PT probe for screening libraries and assessing gene expression patterns

PT Disclosure; Page 40-41; 41pp; English.

XX The present sequence is a cDNA encoding novel human protein (NHP),

CC human protease. NHPs share structural similarity with animal proteases

CC particularly zinc metalloproteases. Sequences of the invention are

CC useful in therapeutic, diagnostic and pharmacogenomic applications.

CC NHP polynucleotides are used as hybridisation probes for screening

CC libraries and assessing gene expression patterns. They can also be

CC used for treating related biological disorders such as obesity, high

CC blood pressure, arthritis, connective tissue disorders and infertility.
XX They are also used in gene therapy.

QY Sequence 3446 BP; 612 A; 1114 C; 1101 G; 619 T; 0 other;

Query Match 99.8%; Score 2846.6; DB 24; Length 3446;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2849; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGCTTTTCTGGGATCCTTAACCTTTCGCGGGGGAACCGCTGGAGGCTTTGAG 60
Db 397 ATGCTTCTCTGGGATCCTTAACCTTTCGCGGGGGAACCGCTGGAGGCTTCTGAG 456

QY 61 CCAGAGCGGGAGTGTAGTTCCTCCACTCGACTGGACCGGACATTAACGGCGCGCTAC 120
Db 457 CCAGAGCGGGAGTGTAGTTCCTCCACTCGACTGGACCGGACATTAACGGCGCGCTAC 516

QY 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCAATTTTCAGATCAACAATTT 180
Db 517 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCAATTTTCAGATCAACAATTT 576

QY 181 CAGGAGGACTTTTACCTACACTGACCGCGGATGCTAGTCTTGGCTCCGCTTCTCC 240
Db 577 CAGGAGGACTTTTACCTACACTGACCGCGGATGCTAGTCTTGGCTCCGCTTCTCC 636

QY 241 ACTGAGCATCTGGGCTCCCTCCAGGGGCTCAACGGGGGCTTTCAGACCTCGACGC 300
Db 637 ACTGAGCATCTGGGCTCCCTCCAGGGGCTCAACGGGGGCTTTCAGACCTCGACGC 696

QY 301 TCGTTCTATTCTGGGAGCTGAACCGCGAGCGGACTCGTTGCTGCTGTGAGCTGTGTC 360
Db 697 TCGTTCTATTCTGGGAGCTGAACCGCGAGCGGACTCGTTGCTGCTGTGAGCTGTGTC 756

QY 361 GGGGGCTCCCGGAGCTTTGGCTACCGAGCGCGGAGTATGTCATTAACCGCTGCCC 420
Db 757 GGGGGCTCCCGGAGCTTTGGCTACCGAGCGCGGAGTATGTCATTAACCGCTGCCC 816

QY 421 AATGCTAGCGCGCGCGGCGAGCAACAGCCAGGCGGACACCTTCTCCAGCGCGG 480
Db 817 AATGCTAGCGCGCGCGGCGAGCAACAGCCAGGCGGACACCTTCTCCAGCGCGG 876

QY 481 GGTGTTCCGGCGCGGCTTCGAGAGACCCACCTCTCGCTGCGGGGTGGCTCGGGTGG 540
Db 877 GGTGTTCCGGCGCGGCTTCGAGAGACCCACCTCTCGCTGCGGGGTGGCTCGGGTGG 936

QY 541 AACCCGCCATCTACGGGCTTGACCCCTTACAGCCGCGCGGCGGGGCTTCGGGAG 600
Db 937 AACCCGCCATCTACGGGCTTGACCCCTTACAGCCGCGCGGCGGGGCTTCGGGAG 996

QY 601 AGTCTAGCGCGCGGCGAGTCTGGGCGCGCAAGCGTTTTCGTCTATCCCGGGTACG 660
Db 997 AGTCTAGCGCGCGGCGAGTCTGGGCGCGCAAGCGTTTTCGTCTATCCCGGGTACG 1056

QY 661 GAGACGCTGGGTGCGGAGCAGTCAATGTTCCAGTTCCAGCGCGGAGCTGGAACAT 720
Db 1057 GAGACGCTGGGTGCGGAGCAGTCAATGTTCCAGTTCCAGCGCGGAGCTGGAACAT 1116

QY 721 TATCTGCTGACGCTGCTGGCAACGCGGAGCTTACCGCCATCCAGCATCCTCAAC 780
Db 1117 TATCTGCTGACGCTGCTGGCAACGCGGAGCTTACCGCCATCCAGCATCCTCAAC 1176

QY 781 CCATCAACATGTTGTTGTTCAAGTGTGCTTCTTAGAGATCGTACTCGGGCCCCAAG 840
Db 1177 CCATCAACATGTTGTTGTTCAAGTGTGCTTCTTAGAGATCGTACTCGGGCCCCAAG 1236

QY 841 GTCACCGCAATGCGGCGCTGACGCTGGCAACTTCTGTGCTGGCAGAAAGAGTGAAC 900
Db 1237 GTCACCGCAATGCGGCGCTGACGCTGGCAACTTCTGTGCTGGCAGAAAGTGAAC 1296

QY 901 AAAGTGTGAGCAAGCAACCCCGAGTCTGGGAGCTGCATCTCTTCCATAGGAGAG 960
Db 1297 AAAGTGTGAGCAAGCAACCCCGAGTCTGGGAGCTGCATCTCTTCCATAGGAGAG 1356

QY 961 CTGTGTGGAGCCACACCTGTGTACACCCCTGGGATGGCTGATGTGGGTACCATGTGTGAC 1020
 Db 1357 CTGTGTGGAGCCACACCTGTGTACACCCCTGGGATGGCTGATGTGGGTACCATGTGTGAC 1416
 QY 1021 CCCAAGAGAAGCTGCTGTCTGTCATTGAGGAGATGGCTTCCATCAGCCCTTACCACTGCGC 1080
 Db 1417 CCCAAGAGAAGCTGCTGTCTGTCATTGAGGAGATGGCTTCCATCAGCCCTTACCACTGCGC 1476
 QY 1081 CAGGAGTGGGCGACGCTGTTCAACATGCCCCATGACATGTAAGTCTGTGAGGAGGTG 1140
 Db 1477 CAGGAGTGGGCGACGCTGTTCAACATGCCCCATGACATGTAAGTCTGTGAGGAGGTG 1536
 QY 1141 TTTGGGAAGCTCGAGCCAAACACATGATGTGTCCTCCCGACCCCTCATCAGATCGACGCTGCC 1200
 Db 1537 TTTGGGAAGCTCGAGCCAAACACATGATGTGTCCTCCCGACCCCTCATCAGATCGACGCTGCC 1596
 QY 1201 AACCCCTGTGTCAGCTGCTGCTGATCATCATCAGCAGCTTCTTGACAGCGGGACGCT 1260
 Db 1597 AACCCCTGTGTCAGCTGCTGCTGATCATCATCAGCAGCTTCTTGACAGCGGGACGCT 1656
 QY 1261 GACTGCTCTCTGGACCAACCCAGCAAGCCCATCTCCCTGCCCGAGGATCTGCCGGGCGCC 1320
 Db 1657 GACTGCTCTCTGGACCAACCCAGCAAGCCCATCTCCCTGCCCGAGGATCTGCCGGGCGCC 1716
 QY 1321 AGCTACACCTGAGCAGCAGTGCAGCTGGCTTTTGGGCTGGGCTCAAGCCCTGTCT 1380
 Db 1717 AGCTACACCTGAGCAGCAGTGCAGCTGGCTTTTGGGCTGGGCTCAAGCCCTGTCT 1776
 QY 1381 TACATGCACTGACACCAAGCTGTGGTGACCGGGAAGCCAAAGGACAGATGTTGTGC 1440
 Db 1777 TACATGCACTGACACCAAGCTGTGGTGACCGGGAAGCCAAAGGACAGATGTTGTGC 1836
 QY 1441 CAGACCGCCCATTTCCCTGGGCGGATGGACACAGCTGTGGGAGGGCAAGCTCTGCCCTC 1500
 Db 1837 CAGACCGCCCATTTCCCTGGGCGGATGGACACAGCTGTGGGAGGGCAAGCTCTGCCCTC 1896
 QY 1501 AAGGGGCTGCTGGAGAGACACCTCAACACACAGGCTGATGGTTCCTGGGCG 1560
 Db 1897 AAGGGGCTGCTGGAGAGACACCTCAACACACAGGCTGATGGTTCCTGGGCG 1956
 QY 1561 AAATGGGATCCCTATGGCCCTCTCGCGACATGTGTGGGGGCTGACGTGGCCAGG 1620
 Db 1957 AAATGGGATCCCTATGGCCCTCTCGCGACATGTGTGGGGGCTGACGTGGCCAGG 2016
 QY 1621 AGGAGTGCACCAACCCACCTGCGACGGGGGCAAGTACTGCGAGGAGTGGGGTG 1680
 Db 2017 AGGAGTGCACCAACCCACCTGCGACGGGGGCAAGTACTGCGAGGAGTGGGGTG 2076
 QY 1681 AAATACCGATCTTGCACCTGGAGCCCTGCCCGCAGCTCAGCCCTCGGAAAGAGCTTCGGG 1740
 Db 2077 AAATACCGATCTTGCACCTGGAGCCCTGCCCGCAGCTCAGCCCTCGGAAAGAGCTTCGGG 2136
 QY 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGCAGCCAAACCGGCTCACTCTCGCC 1800
 Db 2137 GAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGCAGCCAAACCGGCTCACTCTCGCC 2196
 QY 1801 GTGGCATGGGTGCCAAGTACTCCGGGCTGTCTCCCGGGACAAAGTGCATCTCATCTGC 1860
 Db 2197 GTGGCATGGGTGCCAAGTACTCCGGGCTGTCTCCCGGGACAAAGTGCATCTCATCTGC 2256
 QY 1861 CGAGCCAAATGGCACTGGCTACTTCTATGTGCTGGCACCCAAAGTGGTGGACGCGAGCTG 1920
 Db 2257 CGAGCCAAATGGCACTGGCTACTTCTATGTGCTGGCACCCAAAGTGGTGGACGCGAGCTG 2316
 QY 1921 TGCTCTCTGACTCCACCTCCGCTGTGTGTCGAAGGCAAGTGCATCAAGGCTGGCTGTGAT 1980
 Db 2317 TGCTCTCTGACTCCACCTCCGCTGTGTGTCGAAGGCAAGTGCATCAAGGCTGGCTGTGAT 2376
 QY 1981 GGAACCTGGGCTCCGAAGAGATTCGACAAAGTGTGGGGTGTGTGGGGAGACAATAAG 2040
 Db 2377 GGAACCTGGGCTCCGAAGAGATTCGACAAAGTGTGGGGTGTGTGGGGAGACAATAAG 2436
 QY 2041 AGCTGCAAGAAGGTGACTGGACTTCTTCCACCAAGCCCATGCATGCTTACATTTCTGGGTG 2100

Db 2437 AGCTSCAAGAAGGTGACTGGACTCTTCCAAAGCCCATGCTAGCTCAATTTCTGTGGTG 2496
 QY 2101 GCATCCCGCAGCGCCCTCAAGCATCGACATCCGACGCGGTTACAAAGGCTGATC 2160
 Db 2497 GCATCCCGCAGCGCCCTCAAGCATCGACATCCGACGCGGTTACAAAGGCTGATC 2556
 QY 2161 GGGGATGACAACCTACCTGGCTCTGAAGAACAGCCAAAGTACCTGCTCAACGGGAT 2220
 Db 2557 GGGGATGACAACCTACCTGGCTCTGAAGAACAGCCAAAGTACCTGCTCAACGGGAT 2616
 QY 2221 TTCGTGTGTGCGGGTGGAGCGGACCTGTGTGAGGGCAGCTCTGCTCGGCTACAGC 2280
 Db 2617 TTCGTGTGTGCGGGTGGAGCGGACCTGTGTGAGGGCAGCTCTGCTCGGCTACAGC 2676
 QY 2281 GGCACGGGACAGCGGTGAGAGCTGTCAGGCTTCCCGGCCCATCTCTGGAGCCCTGACC 2340
 Db 2677 GGCACGGGACAGCGGTGAGAGCTGTCAGGCTTCCCGGCCCATCTCTGGAGCCCTGACC 2736
 QY 2341 GTGGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTTCTATCTG 2400
 Db 2737 GTGGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTTCTATCTG 2796
 QY 2401 CCCAAGAGCCTCGGAGGACAAAGTCTCTCATCCCAAGGACCCCGGGACCCCTCTGTC 2460
 Db 2797 CCCAAGAGCCTCGGAGGACAAAGTCTCTCATCCCAAGGACCCCGGGACCCCTCTGTC 2856
 QY 2461 TTGCAACAACAGCGTCTCAGCCTCTCCAAACAGGTGGAGCAGCCGACAGAGCCCTCT 2520
 Db 2857 TTGCAACAACAGCGTCTCAGCCTCTCCAAACAGGTGGAGCAGCCGACAGAGCCCTCT 2916
 QY 2521 GCACCTGGGTGGTGGAGCTGGGGGCGTCTCCCGGAGCTGGGAGTGGGCTGCAG 2580
 Db 2917 GCACCTGGGTGGTGGAGCTGGGGGCGTCTCCCGGAGCTGGGAGTGGGCTGCAG 2976
 QY 2581 AAGCGGGCGGTGGACTGCGGGGCTCCCGCGGAGCGCAGCTGCTGCTGTGATGCA 2640
 Db 2977 AAGCGGGCGGTGGACTGCGGGGCTCCCGCGGAGCGCAGCTGCTGCTGTGATGCA 3036
 QY 2641 GCGCATCGCCGTGGAGACAAAGCTGCGGGGAGCCCTGCCACCTGGGAGCTCAGC 2700
 Db 3037 GCGCATCGCCGTGGAGACAAAGCTGCGGGGAGCCCTGCCACCTGGGAGCTCAGC 3096
 QY 2701 GCTGCTCAGCTGCTCCAAAGAGCTGCGGCGGGATTTTCAGAGCGCTCACTCAAGTGT 2760
 Db 3097 GCTGCTCAGCTGCTCCAAAGAGCTGCGGCGGGATTTTCAGAGCGCTCACTCAAGTGT 3156
 QY 2761 GTGGGCCACGGAGCGGCTGCTGGCCCGGACCAAGTGCATTTGCACCGCAAGCCAG 2820
 Db 3157 GTGGGCCACGGAGCGGCTGCTGGCCCGGACCAAGTGCATTTGCACCGCAAGCCAG 3216
 QY 2821 GAGCTGGACTTCTGCGTCTGAGGCCGTGCTGA 2853
 Db 3217 GAGCTGGACTTCTGCGTCTGAGGCCGTGCTGA 3249

RESULT 4

AAS97182

ID AAS97182 standard; cDNA; 2937 BP.

XX AAS97182;

AC AAS97182;

XX 26-FEB-2002 (first entry)

DT 26-FEB-2002 (first entry)

XX Human metalloprotease partial DNA sequence #11.

DE Human metalloprotease partial DNA sequence #11.

XX Human metalloprotease partial DNA sequence #11.

XX Human metalloprotease partial DNA sequence #11.

XX Human metalloprotease partial DNA sequence #11.

XX Human metalloprotease partial DNA sequence #11.

XX Human metalloprotease partial DNA sequence #11.

XX Human metalloprotease partial DNA sequence #11.

XX Human metalloprotease partial DNA sequence #11.

XX Human metalloprotease partial DNA sequence #11.

XX Human metalloprotease partial DNA sequence #11.

XX Human metalloprotease partial DNA sequence #11.

XX Human metalloprotease partial DNA sequence #11.

XX Human metalloprotease partial DNA sequence #11.

XX Human metalloprotease partial DNA sequence #11.

Human metalloprotease partial DNA sequence #11.
 Human: protease; PCR primer; cytosolic; immunomodulator, cardiac;
 vasotropic; antinigraine; analgesic; endocrine; norepinephrine; anabolic;
 hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;
 anorectic; antiinflammatory; aspartyl protease; cysteine protease;
 metalloprotease; serine protease; cancer; haematopoietic; breast;
 lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;
 immune-related disease; cardiovascular disease; neuronal disease;

QY	1441	CAGACCCGCCACTTCCCCTGGCCCGCATGGCACACCAAGCTGTGGCAGGGCAAGCTCTCGCTC	1500
Db	1516	CAGACCCGCCACTTCCCCTGGCCCGCATGGCACACCAAGCTGTGGCAGGGCAAGCTCTCGCTC	1575
QY	1501	AAAGGGCCCTGGTGGAGAGACACAACCTCAACAAAGACAAAGGTGGATGTTCTCTGGGC	1560
Db	1576	AAAGGGCCCTGGTGGAGAGACACAACCTCAACAAAGACAAAGGTGGATGTTCTCTGGGC	1635
QY	1561	AAATGGATCCCTATGAGCCCTGCTCGCGCACATGTGGTGGGGCGTGCAGCTGGCCAGG	1620
Db	1636	AAATGGATCCCTATGAGCCCTGCTCGCGCACATGTGGTGGGGCGTGCAGCTGGCCAGG	1695
QY	1621	AGGCAGTGACCAACCCCAACCCCTGCCAACCGGGGCAAGTACTGCGAGGGAGTGAAGGTG	1680
Db	1696	AGGCAGTGACCAACCCCAACCCCTGCCAACCGGGGCAAGTACTGCGAGGGAGTGAAGGTG	1755
QY	1681	AAATACCGATCCTGCAACCTGGAGCCCTGCCAGCTCAGCCTCCGGAAAGAGCTTCCGG	1740
Db	1756	AAATACCGATCCTGCAACCTGGAGCCCTGCCAGCTCAGCCTCCGGAAAGAGCTTCCGG	1815
QY	1741	GAGGAGCAGTGTGAGGCTTTCAACGGCTTAAACACACAGCACCAACCGGCTCACTCTCGCC	1800
Db	1816	GAGGAGCAGTGTGAGGCTTTCAACGGCTTAAACACACAGCACCAACCGGCTCACTCTCGCC	1875
QY	1801	GTGGCATGGGTGCCCAAGTACTTCGGCGCTGTCTCCCGGGACAAGTCAAGCTCATCTGC	1860
Db	1876	GTGGCATGGGTGCCCAAGTACTTCGGCGCTGTCTCCCGGGACAAGTCAAGCTCATCTGC	1935
QY	1861	CGAGCCAATGGCAGCTGGCTACTTCTATGTCTGGCACCACCA -- --GGTGGTGGAGCGCAG	1917
Db	1936	CGAGCCAATGGCAGCTGGCTACTTCTATGTCTGGCACCACCAAGTGGTGGAGCGGCAG	1995
QY	1918	CTGTGCTCTCTGACTTCCACCTCCGCTGTGTCTCAAGGCAAGTGCATCAAGGCTGGCTGT	1977
Db	1996	CTGTGCTCTCTGACTTCCACCTCCGCTGTGTCTCAAGGCAAGTGCATCAAGGCTGGCTGT	2055
QY	1978	GATGGAAACCTGGGCTTCCAAGAAAGATTTGCACAAGTGTGGGGTGTGTGGGGGAGACAAT	2037
Db	2056	GATGGAAACCTGGGCTTCCAAGAAAGATTTGCACAAGTGTGGGGTGTGTGGGGGAGACAAT	2115
QY	2038	AAGAGCTGCAAGAAGTGACTGAGCTCTTCAACAGCCCATGCATGCTACAAATTTCTGTG	2097
Db	2116	AAGAGCTGCAAGAAGTGACTGAGCTCTTCAACAGCCCATGCATGCTACAAATTTCTGTG	2175
QY	2098	GTGGCCATCCCGCAGCGCCTCAAGCATCGACATCCGCCAGCGCGTTCACAAGGGCTG	2157
Db	2176	GTGGCCATCCCGCAGCGCCTCAAGCATCGACATCCGCCAGCGCGTTCACAAGGGCTG	2235
QY	2158	ATCGGGGATGACAATCTACCTGGCTCTGAAGAACAGCCAAAGCAAGTACCTGCTCAACGGG	2217
Db	2236	ATCGGGGATGACAATCTACCTGGCTCTGAAGAACAGCCAAAGCAAGTACCTGCTCAACGGG	2295
QY	2218	CATTTCTGTGTGCGCGTGGAGCGGACCTGTGTGTGAAGGCGAGTGTCTGCTGCGGTAC	2277
Db	2296	CATTTCTGTGTGCGCGTGGAGCGGACCTGTGTGTGAAGGCGAGTGTCTGCTGCGGTAC	2355
QY	2278	AGCGGCACGGGCACAGCGTGGAGAGCCTGCAAGGCTTCCCGGCCCATCTCTGGAGCCGCTG	2337
Db	2356	AGCGGCACGGGCACAGCGTGGAGAGCCTGCAAGGCTTCCCGGCCCATCTCTGGAGCCGCTG	2415
QY	2338	ACCGTGGAGTCTCTCCGTGGGGAGATGACACCGCCCCCGGTTCGGGTACTCTTCTAT	2397
Db	2416	ACCGTGGAGTCTCTCCGTGGGGAGATGACACCGCCCCCGGTTCGGGTACTCTTCTAT	2475
QY	2398	CTGCCAAMAGAGCTCGGGGAGACAAGTCTCTCATCCCAAGGACCCCGG -- --GGA	2451
Db	2476	CTGCCAAMAGAGCTCGGGGAGACAAGTCTCTCATCCCGGACCCCGGGGAGGAGA	2535
QY	2452	CCCTCTGTCTTGCACAACAGCTCTCTAGCCTCTCCAAACCAAGTGGAGACGCGGACGAC	2511
Db	2536	CCCTCTGTCTTGCACAACAGCTCTCTAGCCTCTCCAAACCAAGTGGAGACGCGGACGAC	2595

Qy	2512	AGGCCCTTCACGCTGGGTGGCTGGAGCTGGGGGCGGTGCTCCCGGAGCTGCGGCAGT	2571
Db	2596	AGGCCCTTCACGCTGGGTGGCTGGAGCTGGGGGCGGTGCTCCCGGAGCTGCGGCAGT	2655
Qy	2572	GGCTGCAGAAAGGGGGGTGGACTGCGGGGGCTCCGGGGGAGCGCACGGTCCCTGCGC	2631
Db	2656	GGCTGCAGAAAGGGGGGTGGACTGCGGGGGCTCCGGGGGAGCGCACGGTCCCTGCGC	2715
Qy	2632	TGTGATCAGACCCATCGGCCCTGGAGACACAAAGCTGCGGGGAGCCCTGCCCCACCTGG	2691
Db	2716	TGTGATCAGACCCATCGGCCCTGGAGACACAAAGCTGCGGGGAGCCCTGCCCCACCTGG	2775
Qy	2692	GAGCTCAGCGCTGGTCACCTGCTCCAAGAGCTGCGGGCGGGGATTTTCAGAGGCGCTCA	2751
Db	2776	GAGCTCAGCGCTGGTCACCTGCTCCAAGAGCTGCGGGCGGGGATTTTCAGAGGCGCTCA	2835
Qy	2752	CTCAAGTGTGTGGGCCACGAGGCCGCGCTGCTGGCCCCGGGACACAGTGCACCTTGACCGC	2811
Db	2836	CTCAAGTGTGTGGGCCACGAGGCCGCGCTGCTGGCCCCGGGACACAGTGCACCTTGACCGC	2895
Qy	2812	AAGCCCCAGAGCTGGACTTCTGGGTCTCAGGCCGTGCTGA	2853
Db	2896	AAGCCCCAGAGCTGGACTTCTGGGTCTCAGGCCGTGCTGA	2937
RESULT 5			
ABK12894			
ID	ABK12894 standard; cDNA; 2930 BP.		
AC	ABK12894;		
XX			
DT	09-APR-2002 (first entry)		
XX			
DE	Human protease PRTS-11 cDNA sequence.		
XX			
KW	Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;		
KW	cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;		
KW	inflammatory disorder; acquired immunodeficiency syndrome; AIDS;		
KW	cell proliferative disorder; developmental disorder; epilepsy;		
KW	Duchenne muscular dystrophy; epithelial disorder; neurological disorder;		
KW	reproductive disorder; endometriosis; ss.		
OS	Homo sapiens.		
XX			
FH	Key Location/Qualifiers		
FT	CDS 75..2930		
FT	/*tag= a		
FT	/partial		
FT	/product= "Human protease PRTS-11"		
FT	/note= "This sequence lacks a stop codon"		
XX			
PN	WO200198468-A2.		
XX			
PD	27-DEC-2001.		
XX			
PF	13-JUN-2001; 2001WO-US19178.		
XX			
PR	16-JUN-2000; 2000US-212336P.		
PR	22-JUN-2000; 2000US-213955P.		
PR	29-JUN-2000; 2000US-215396P.		
PR	07-JUL-2000; 2000US-216821P.		
PR	14-JUL-2000; 2000US-218946P.		
XX			
PA	(INCY-) INCYTE GENOMICS INC.		
XX			
PI	Yue H, Elliott VS, Gandhi AR, Lal P, Au-Young J, Tribouley CM;		
PI	Delegane AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;		
PI	Walja NK, Yao MG, Lu DM, Patterson C, Tang YT, Walsh RT;		
PI	Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;		
PI	Kallick DA;		
XX			
DR	WPI; 2002-090437/12.		
DR	P-PSDB; RAU74751.		

Db	1815	GAGGACAGTGTGAGGCTTTCAACGGCTTACAACCAACAGCACAGCACCACCGGCTCACTCTCGCC	1871
Qy	1801	GTGGCATGGGTGCCCAAGTACTCCGGCGTGTCTCCCGGGACAAAGTCAAGCTCATCTGC	1860
Db	1875	GTGGCATGGGTGCCCAAGTACTCCGGCGTGTCTCCCGGGACAAAGTCAAGCTCATCTGC	1934
Qy	1861	CGAGCCATGGGCACGTGGCTACTTTATGTGCTGGCACCCCA--GGTGGTGGAGCGCAG	1917
Db	1935	CGAGCCATGGGCACGTGGCTACTTTATGTGCTGGCACCCCAAGTGGTGGAGCGCAG	1994
Qy	1918	CTGTGCTCTCCTGACTCCACCTCCGCTGTGTCTCAAGGCAAGTGCATCAAGGCTGGCTGT	1977
Db	1995	CTGTGCTCTCCTGACTCCACCTCCGCTGTGTCTCAAGGCAAGTGCATCAAGGCTGGCTGT	2054
Qy	1978	GATGGAACTCGGGCTCCAAAGAAGAGATTGCACAAAGTGTGGGTGTGTGGGGAGACAAT	2037
Db	2055	GATGGAACTCGGGCTCCAAAGAAGAGATTGCACAAAGTGTGGGTGTGTGGGGAGACAAT	2114
Qy	2038	AAGAGCTGCAAGAAGTGACTGGACTCTTACCAGGCCATGCAATGCTACAATTCGTG	2097
Db	2115	AAGAGCTGCAAGAAGTGACTGGACTCTTACCAGGCCATGCAATGCTACAATTCGTG	2174
Qy	2098	GTGGCCATCCCGAGCGCCTCAAGCATCGACATCCGCCAGCGCGTTACAAGGCGTG	2157
Db	2175	GTGGCCATCCCGAGCGCCTCAAGCATCGACATCCGCCAGCGCGTTACAAGGCGTG	2234
Qy	2158	ATCGGGATGACACTACCTGCTGTGAAGACACGCCAAGCAAGTACTGCTCAACGG	2217
Db	2235	ATCGGGATGACACTACCTGCTGTGAAGACACGCCAAGCAAGTACTGCTCAACGG	2294
Qy	2218	CATTTCGTGTGTGCGCGGTGGAGCGGACCTGTGTGTGAAGGCGAGTCTGCTCGGTAC	2277
Db	2295	CATTTCGTGTGTGCGCGGTGGAGCGGACCTGTGTGTGAAGGCGAGTCTGCTCGGTAC	2354
Qy	2278	AGCGGCACGGGCACAGCGGTGAGAGCCTGCAGGCTTCCCGGCCCATCTCTGGAGCGCGTG	2337
Db	2355	AGCGGCACGGGCACAGCGGTGAGAGCCTGCAGGCTTCCCGGCCCATCTCTGGAGCGCGTG	2414
Qy	2338	ACCGTGGAGGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTCTCTAT	2397
Db	2415	ACCGTGGAGGTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTCTCTAT	2474
Qy	2398	CTGCCAAGAGCGCTCGGAGGACAGTCTCTCATNCCCAAGACCCCGG---GGACCC	2454
Db	2475	CTGCCAAGAGCGCTCGGAGGACAGTCTCTCATNCCCGCGCACCCCGGGGAGGACCC	2534
Qy	2455	TCTGTCTTGACACAACAGCGTCTCAGGCTCTCCAACAGGTGGAGCAGCGGACGACAGG	2514
Db	2535	TCTGTCTTGACACAACAGCGTCTCAGGCTCTCCAACAGGTGGAGCAGCGGAGCAGG	2594
Qy	2515	CCCCCTGCAGCTGGGTGGCTGGCAGCTGGGGCGGTCTCCGCGAGCTGCGGCAGTGGC	2574
Db	2595	CCCCCTGCAGCTGGGTGGCTGGCAGCTGGGGCGGTCTCCGCGAGCTGCGGCAGTGGC	2654
Qy	2575	CTGCAGAGCGCGGTGACCTGCGGGGCTCCCGCGGCGAGCGCAGGTCCTGCGCTGT	2634
Db	2655	CTGCAGAGCGCGGTGACCTGCGGGGCTCCCGCGGCGAGCGCAGGTCCTGCGCTGT	2714
Qy	2635	GATGCAGCCCATCGGCCGTGGAGACACAAGCCTTCGGGGAGCCCTGCCACCTGGGAG	2694
Db	2715	GATGCAGCCCATCGGCCGTGGAGACACAAGCCTTCGGGGAGCCCTGCCACCTGGGAG	2774
Qy	2695	CTCAGCGCTGTACCTGTCTCAAGAGCTGCGGCCCGGGGATTTTCAGAGCGCTCACTC	2754
Db	2775	CTCAGCGCTGTGTACCTGTCTCAAGAGCTGCGGCCCGGGGATTTTCAGAGCGCTCACTC	2834
Qy	2755	AAGTGTGTGGCCACGAGGCGCGTGTGTGGCCCGGGACACAGTGCACCTTCACCCGAAG	2814
Db	2835	AAGTGTGTGGCCACGAGGCGCGTGTGTGGCCCGGGACACAGTGCACCTTCACCCGAAG	2894
Qy	2815	CCCCAGGAGCTGACTTCTCGCTCTGAGGCGCGTGC	2850
Db	2895	CCCCAGGAGCTGACTTCTCGCTCTGAGGCGCGTGC	2930

Db 97 GCGAGCCAGACTCCTTTGCTGCTAAGCTATGCGGGGTCTCCGCGAGCCTTTGGC 156
Qy 385 TACGAGCGCGAGTATGATAGCCGCTGCCAATGCTAGCGCGCGCGGCGCAG 444
Db 157 TACCAAGGTGCGAGTATGATAGCCCTTGTGCCAACACACGCGCGCTGAGGCGAG 216
Qy 445 CGCAACAGCCAGGCGCACACCTTCTCCAGCGCGGGGTGTTCCGGCGCGCCCTCCGGA 504
Db 217 CGTCATAGCCAGGCGCACACCTTCTCCAGCGCGGGGTGCTCCCGTAGGCGCTTCCGGA 276
Qy 505 GACCCCACTCTCGCTCGGGGTGCGCTCGGGGTGGAACCCCGCATCTTACGCGGCGCTG 564
Db 277 GACCTTACCTCTCGCTCGGGGTGCGCTCGGGGTGGAACCCCGCATCTTACGCGGCTG 336
Qy 565 GACCTTTACAAGCGCGCGGGGTGCTCGGGGTGAGAGTCTGAGCGCGCGAGTCTGG 624
Db 337 GACCTTTAACAACAGCGCGGGGTGCGGGGTGAAGCCACAAACCGCGCGAGTCTGG 336
Qy 625 CGCGCAAGCGTTTCTGCTATCCCGCGGTACGTGGAGACGCTGGTGGCGGCGAG 684
Db 397 CGCGCAAGCGCTTCTGCTATACCAACGCTAGCTGGAGACACTGGTGGCGGAGAG 456
Qy 685 TCAATGCTCAAGTTTCCAGCGCGGAGCTGGAACATATCTGCTGAGCTGCTGGCAAG 744
Db 457 TCAATGCTCAAGTTTCCAGCGCGGATTTGGAACATATCTGCTGAGCTGCTGGCAG 516
Qy 745 GCGCGGAGCTTACCGCCATCCAGCATCTCAACCCCATCAACATCTGTTGGTCAAG 804
Db 517 GCGCGGAGCTTACCGCCATCCAGCATCTCAACCCCATCAACATCTGTTGGTCAAG 576
Qy 805 GTGCTGCTTTAGAGATCTGACTCGGGGCCAAGTCAACCGCAATGCGGCCCTGAGC 864
Db 577 GTGTTACTCTTAGAGATCTGACTCGGGGCCAAGTCAACCGCAATGCGGCCCTGACT 636
Qy 865 CTGCGAATCTTGTGCTGCGGAGAGAGCTCAACAAAGTGAAGTCAACGACACCCGAG 924
Db 637 CTGCGAATCTTGTGCTGCGGAGAGAGTGAACAAAGTGAAGTCAACGACACCCGAG 696
Qy 925 TACTGGAGACTGCTGCTTCAACAGCGAGGAGCTGTGGAGCCACACCTGTGAC 984
Db 697 TACTGGAGACTGCTGCTTCAACAGCGAGGAGCTGTGGAGCCACACCTGTGAC 756
Qy 985 ACCTGGGATGCTGATGTTGCTGCTGACCTGCGGCGGAGAGTCTGCTGTCTT 1044
Db 757 ACCTGGGATGCTGATGTTGCTGCTGACCTGCGGCGGAGAGTCTGCTGTCTT 816
Qy 1045 GAGGAGATGGGCTTCCATCAGCTTCAACAGCGAGGAGTGTGGAGCTCCGAGCCAAACC 1104
Db 817 GAGGAGATGGGCTTCCATCAGCTTCAACAGCGAGGAGTGTGGAGCTCCGAGCCAAACC 876
Qy 1105 ATGCCCATGACAATGTGAAGTCTGTGAGGAGTGTGGAGCTCCGAGCCAAACC 1164
Db 877 ATGCCCATGACAATGTGAAGTCTGTGAGGAGTGTGGAGCTCCGAGCCAAACC 936
Qy 1165 ATGATGTCCCGACCTCTACAGATCGACCTGCGCAACCCCTGTGACCTGCGAGTGT 1224
Db 937 ATGATGTCTCCGACATCTACAGATCGACCTGCGCAACCCCTGTGACCTGCGAGTGT 996
Qy 1225 GCATCATCACCAGCTTCTTGGACAGCGGCGAGCTGTGCTGCTGCTGAGCCAAACC 1284
Db 997 GCATCATCACCAGCTTCTTGGACAGCGGCGAGCTGTGCTGCTGCTGAGCCAAACC 1056
Qy 1285 AAGCCCATCTCCCTGCGGAGATCTGCGGCGGAGTACACCTGAGCCAGAGTGC 1344
Db 1057 AAGCCCATCTCCCTGCGGAGATCTGCGGCGGAGTACACCTGAGCCAGAGTGC 1116
Qy 1345 GAGCTGGCTTTGGGCTGGGCTCCAGCCCTGCTCTTACATGCTGCTGCTGCTGCTGCTG 1404
Db 1117 GAGCTGGCTTTGGGCTGGGCTCCAGCCCTGCTCTTACATGCTGCTGCTGCTGCTGCTG 1176
Qy 1405 TGSTGACCGGAGGAGGAGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1464
Db 1177 TGSTGACCGGAGGAGGAGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1236

Qy 1465 GATGGCACAGCTGTGGCGAGGCAAGCTCTCCTCAAGGGGCTGCTGCTGAGAGACAC 1524
Db 1237 GATGGCACAGCTGTGGCGAGGCAAGCTCTCCTCAAGGGGCTGCTGCTGAGAGACAC 1296
Qy 1525 AACCTCAACAACAGCAGGCTGATGTTCTCTGGGCAATGGGATCCCTATGGCCCTGTC 1584
Db 1297 AACCCAAACAGTACCGGCTGGAGCGCCCTTGGGCAAGTGGGAGCCCTATATGGTCCCTGC 1356
Qy 1585 TCGCGCACATGTTGGTGGGCGCTGACGCTGGCGAGGAGCA--GTGCACCAACCCACCC 1642
Db 1357 TCGCGCACCTGCTGGTGGGCGCGCAGCTGGCCCGGAGGCAAGTCAAGCAACCTATACC 1416
Qy 1643 CTGCCAACGGG--GGCAAGTACTGCGAGGAGTGAAGTGAATACCGATCTCTGCAACCTG 1701
Db 1417 CTGCCAACGGGCGGGAAGTACTGCGAGGAGTGAAGTGAATACCGATCTCTGCAACCTG 1476
Qy 1702 GAGCCCTGCGCCAGCTCAGCCTCCGAAAGAGCTTCCGCGGA 1742
Db 1477 GAGCCCTGCGCCAGCTCAGCCTCTGCAAGAGCTTCCGCGGA 1517

RESULT 7
AAD35570

ID AAD35570 standard; cDNA; 1104 BP.

XX AAD35570;

XX AC AAD35570;

XX DT 26-JUL-2002 (first entry)

XX DE Human protease cDNA #3.

XX KW Human; novel human protein; NHP; protease; biological disorder; obesity;
XX KW high blood pressure; arthritis; connective tissue disorder; infertility;
XX KW gene therapy; enzyme; gene; ss.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers
XX CDS 1..1104XX FT /*tag= a
XX FT /product= "Human protease #3"

XX PN WO200226949-A2.

XX PD 04-APR-2002.

XX PF 27-SEP-2001; 2001WO-US30350.

XX PR 29-SEP-2000; 2000US-236689P.

XX PA (LEXI-) LEXICON GENETICS INC.

XX PI Friddle CJ, Hilbun E;

XX DR WPI; 2002-372123/40.

XX DR P-PSDB; AAE22542.

XX PT Novel nucleic acid encoding a human protease, useful as a hybridization
XX PT probe for screening libraries and assessing gene expression patterns -
XX FS Disclosure; Page 38-39; 41pp; English.XX CC The present sequence is a cDNA encoding novel human protein (NHP),
XX CC human protease. NHPs share structural similarity with animal proteases
XX CC particularly zinc metalloproteases. Sequences of the invention are
XX CC useful in therapeutic, diagnostic and pharmacogenomic applications.
XX CC NHP polynucleotides are used as hybridisation probes for screening
XX CC libraries and assessing gene expression patterns. They can also be
XX CC used for treating related biological disorders such as obesity, high
XX CC blood pressure, arthritis, connective tissue disorders and infertility.
XX CC They are also used in gene therapy.

XX

Query Match 36.6%; Score 1045.2; DB 21; Length 1143;
Best Local Similarity 96.6%; Pred. No. 1.le-207;

Matches 1085; Conservative 2; Mismatches 33; Indels 3; Gaps 2;

QY 738 GGCAAGCGCGGAGCTCTACCGCATCCAGCATCCCTCAACCCCATCAACATCTTGT 797
Db 21 GGCGCCCGGCGAGTATCTTTAGCATCCAGCATCCCTCAACCCCATCAACATCTTGT 80

QY 798 GGTCAGGTGCTGCTTTAGAGATCGTGACTCCGGGCCCAAGTCCACGGCAATCGGC 857
Db 81 GGTCAGGTGCTGCTTTAGAGATCGTGACTCCGGGCCCAAGTCCACGGCAATCGGC 140

QY 858 CCTGACGCTCGCAACTTCTGTCCTGGCAGAGAGAGCTGAACAAAGTAGTGACAAGCA 917
Db 141 CCTGACGCTCGCAACTTCTGTCCTGGCAGAGAGAGCTGAACAAAGTAGTGACAAGCA 200

QY 918 CCCGAGTACTGGGACACTGCCATCTCTTCCAGGAGGAGCTGTGGAGCCACCAC 977
Db 201 CCCGAGTACTGGGACACTGCCATCTCTTCCAGGAGGAGCTGTGGAGCCACCAC 260

QY 978 CTGTGACACCTCGGCATGCTGATGGGTACCATGTGTGACCCCAAGAGAGCTGCTC 1037
Db 261 CTGTGACACCTCGGCATGCTGATGGGTACCATGTGTGACCCCAAGAGAGCTGCTC 320

QY 1038 TGTGATGAGGAGATGGGCTTCCATCAGCTTCACACTGCCAGAGCTGGGCCACGT 1097
Db 321 TGTGATGAGGAGATGGGCTTCCATCAGCTTCACACTGCCAGAGCTGGGCCACGT 380

QY 1098 GTTCAACATGCCCATGACAAATGTGAGAGGCTGTTGGGAAGCTCGGAGC 1157
Db 381 GTTCAACATGCCCATGACAAATGTGAGAGGCTGTTGGGAAGCTCGGAGC 440

QY 1158 CAACCATGATGTCCCGGACCTCATCCAGATCGACCTGCGCAACCCCTGTGACCCG 1217
Db 441 CAACCATGATGTCCCGGACCTCATCCAGATCGACCTGCGCAACCCCTGTGACCCG 500

QY 1218 CAGTGCTGCTCATCAGGACCTTCTGACAGCGGAGCTGACCTGCTCTGAGACA 1277
Db 501 CAGTGCTGCTCATCAGGACCTTCTGACAGCGGAGCTGACCTGCTCTGAGACA 560

QY 1278 ACCAGCAAGCCCATCTCCCTGCGGAGGATCTGCGGGGCGGAGCTACACCTGAGCA 1337
Db 561 ACCAGCAAGCCCATCTCCCTGCGGAGGATCTGCGGGGCGGAGCTACACCTGAGCA 620

QY 1338 GCAGTGGAGCTGGCTTTGGCTGGGCTCCAAAGCCCTGCTTACATGAGTACTGCAC 1397
Db 621 GCAGTGGAGCTGGCTTTGGCTGGGCTCCAAAGCCCTGCTTACATGAGTACTGCAC 680

QY 1398 CAAAGCTGTGTGACCGGAGGAGGAGGAGATGCTGCCAGCCGCGGAGCTTCC 1457
Db 681 CAAAGCTGTGTGACCGGAGGAGGAGGAGATGCTGCCAGCCGCGGAGCTTCC 740

QY 1458 CTGGGCGGATGGCACCAGCTGTGGCGAGGGCAAGCTCTGCCCTCAAGGGGCGCTGGGA 1517
Db 741 CTGGGCGGATGGCACCAGCTGTGGCGAGGGCAAGCTCTGCCCTCAAGGGGCGCTGGGA 800

QY 1518 GAGACACACCTCAACAGCAGGAGTGTGCTTCTGGGCAATGGATGCCCTATGG 1577
Db 801 AARACACACCTCAACAGCAGGAGTGTGCTTCTGGGCAATGGATGCCCTATGG 860

QY 1578 CCCTGTCTGCGCACATGTGTGGGGCGTGCAGCTGGCGGAGGAGGAG-TCGACCAACC 1636
Db 861 CCCTGTCTGCGCACATGTGTGGGGCGTGCAGCTGGCGGAGGAGGAGTGCACCAACC 920

QY 1637 CCACCCC--TGCCAACGGGGCAAGTACTGCGAGGAGTGTGAGGTTGAATACCGATCCTG 1694
Db 921 CCANCCCTGCCAACNGGGGGCAAGTACTGCGAGGAGTGTGAGGTTGAATACCGATCCTG 980

QY 1695 CAACCTGGAGCCCTGCCAGCTCAGCTCCGGAAGAGCTTCGCGGAGGAGGAGTGTGA 1754
Db 981 CAACCTGGAGCCCTGCCAGCTCAGCTCCGGAAGAGCTTCGCGGAGGAGGAGTGTGA 1040

QY 1755 GGCTTTCAACGGGTACAAACACAGCAGCAACAGGCTCTCTCGCGCTGGCATGGTGGC 1814
Db 1041 GGCTTTCAACGGGTACAAACACAGCAGCAACAGGCTCTCTCGCGCTGGCATGGTGGC 1100

QY 1815 CAACTACTCCGCGTGTCTCCCGGAGCAAGTGCAGCTCATC 1857
Db 1101 CAACTACTCCGCGTGTCTCCCGGAGCAAGTGTAACTCATC 1143

RESULT 9
AAD35568
ID AAD35568 standard; cDNA; 966 BP.
XX AAD35568;
AC AAD35568;
XX
DT 26-JUL-2002 (first entry)
XX
DE Human protease cDNA #1.
XX
KW Human; novel human protein; NHP; protease; biological disorder; obesity;
high blood pressure; arthritis; connective tissue disorder; infertility;
gene therapy; enzyme; gene; ss.
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
CDS 1..966
FT /*tag= a
FT /product= "Human protease #1"
XX
PN WO200226949-A2.
XX
PD 04-APR-2002.
XX
PF 27-SEP-2001; 2001WO-US30350.
XX
PR 29-SEP-2000; 2000US-236689P.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Friddle CJ, Hilbun E;
XX
DR WPI: 2002-372123/40.
DR P-PSDB: AAE22540.
XX
XX Novel nucleic acid encoding a human protease, useful as a hybridization probe for screening libraries and assessing gene expression patterns -
PS Disclosure; Page 34; 41pp; English.
XX
CC The present sequence is a cDNA encoding novel human protein (NHP),
human protease. NHPs share structural similarity with animal proteases
particularly zinc metalloproteases. Sequences of the invention are
useful in therapeutic, diagnostic and pharmacogenomic applications.
CC NHP polynucleotides are used as hybridisation probes for screening
libraries and assessing gene expression patterns. They can also be
used for treating related biological disorders such as obesity, high
blood pressure, arthritis, connective tissue disorders and infertility.
CC They are also used in gene therapy.
XX
SQ Sequence 966 BP; 161 A; 316 C; 305 G; 184 T; 0 other;

Query Match 33.5%; Score 955.8; DB 24; Length 966;
Best Local Similarity 99.8%; Pred. No. 3.8e-189;
Matches 957; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCTTTGCTGGCATCTTAACCTGGCTTTCGCGGCAACCGCTGGAGGCTTTCAG 60
Db 1 ATGCTTTGCTGGCATCTTAACCTGGCTTTCGCGGCAACCGCTGGAGGCTTTCAG 60

QY 61 CCAGAGCGGAGGTAGTGTGCTTCCATCGACTGGACCGGACATTAACGGCGCGCTAC 120
Db 61 CCAGAGCGGAGGTAGTGTGCTTCCATCGACTGGACCGGACATTAACGGCGCGCTAC 120

QY 121 TACTGGCGGGTCCGAGGACTCCGGGATCCAGGACTCATTTTTCAGATCACACATTT 180
Db 121 TACTGGCGGGTCCGAGGACTCCGGGATCCAGGACTCATTTTTCAGATCACACATTT 180

Db 121 TACTGGCGGGTCCCGAGGACTCCGGGGATCAGGGACTCAATTTTCAGATCAGACATTT 180
 QY 181 CAGGAGACTTTTACCTACACCTGACCGGATGCTCACTTCTTGCTCCCGCTTCTCC 240
 Db 181 CAGGAGACTTTTACCTACACCTGACCGGATGCTCACTTCTTGCTCCCGCTTCTCC 240
 QY 241 ACTGAGCATCTGGGCTCCCTCCAGGGGCTCACGGGGGCTTTCAGACCTGCGAGC 300
 Db 241 ACTGAGCATCTGGGCTCCCTCCAGGGGCTCACGGGGGCTTTCAGACCTGCGAGC 300
 QY 301 TGCTTCTATCTGGGAGCTGAACCGCGGAGCTGTTGCTGCTGTGAGCTGTGC 360
 Db 301 TGCTTCTATCTGGGAGCTGAACCGCGGAGCTGTTGCTGCTGTGAGCTGTGC 360
 QY 361 GGGGGGCTCCGGAGCTTTGGCTACCGAGCGCGAGTATGTATAGCCCGCTGCCC 420
 Db 361 GGGGGGCTCCGGAGCTTTGGCTACCGAGCGCGAGTATGTATAGCCCGCTGCCC 420
 QY 421 AATGCTAGGCGCGCGGCGGAGCGCAACAGCGAGGCGCACCTTCTCAGCGCGCG 480
 Db 421 AATGCTAGGCGCGCGGCGGAGCGCAACAGCGAGGCGCACCTTCTCAGCGCGCG 480
 QY 481 GGTGTTCCGGCGGCGCTTCCGGAGACCCACCTCTCGCTGCGGGTGGCTCGGCTGG 540
 Db 481 GGTGTTCCGGCGGCGCTTCCGGAGACCCACCTCTCGCTGCGGGTGGCTCGGCTGG 540
 QY 541 AACCCGCCATCTACGGGCGCTTGGACCTTACAGCCGCGGCGGCGGCTTCGGGGAG 600
 Db 541 AACCCGCCATCTACGGGCGCTTGGACCTTACAGCCGCGGCGGCGGCTTCGGGGAG 600
 QY 601 AGTCGTAGCGCGGAGGCTTGGGCGCGCGAGCGTTCGTGTCTATCCCGGCTACGTG 660
 Db 601 AGTCGTAGCGCGGAGGCTTGGGCGCGCGAGCGTTCGTGTCTATCCCGGCTACGTG 660
 QY 661 GAGACGCTGTGTGCGGAGAGTCAATGGTCAAGTTCACGGCGCGACCTTGAACAT 720
 Db 661 GAGACGCTGTGTGCGGAGAGTCAATGGTCAAGTTCACGGCGCGACCTTGAACAT 720
 QY 721 TATCTGCTACGCTGTGGCAAGCGGCGGAGCTTACCGCATCCAGCATCTCAAC 780
 Db 721 TATCTGCTACGCTGTGGCAAGCGGCGGAGCTTACCGCATCCAGCATCTCAAC 780
 QY 781 CCATCAACATCTGTGGTCAAGGTGCTCTTCTTAGAGATGCTGACTCCGGGCGCAAG 840
 Db 781 CCATCAACATCTGTGGTCAAGGTGCTCTTCTTAGAGATGCTGACTCCGGGCGCAAG 840
 QY 841 GTACCGGCAATCGGCGCTGACGCTCGCAACTTCTGCTGCGGAGAGAGCTGAAC 900
 Db 841 GTACCGGCAATCGGCGCTGACGCTCGCAACTTCTGCTGCGGAGAGAGCTGAAC 900
 QY 901 AAAGTGAGTGACAGCAACCGGAGTACTGGGACACTGCCATCTCTTACCGAGGAGGA 959
 Db 901 AAAGTGAGTGACAGCAACCGGAGTACTGGGACACTGCCATCTCTTACCGAGGAGGA 959

RESULT 10

AAH20226
 ID AAH20226 standard; DNA: 2670 BP.

XX AC AAH20226;

XX DX 07-AUG-2001 (first entry)

XX DE Human ADAM-type metalloprotease-related probe, SEQ ID NO:3.

XX DE Human: MDT54; ADAM-type metalloprotease; drug screening;

XX KW A Disintegrin And Metalloprotease; cancer; arthritis; probe; ss.

XX OS Homo sapiens.

XX XX JP2001017183-A.

XX XX 23-JAN-2001.

XX 09-JUL-1999; 99JP-0196584.
 XX 09-JUL-1999; 99JP-0196584.
 XX (YAMA) YAMANOUCHI PHARM CO LTD.
 XX WPI; 2001-275950/29.
 XX A new metal protease and its preparation for use as an anti-cancer and
 XX anti-arthritis therapeutic -
 XX Example 1; Page 11-12; 22pp; Japanese.
 XX The invention relates to the novel human ADAM (A Disintegrin And
 XX Metalloprotease)-type metalloproteases MDT54 (AAB73549) and MDT55
 XX (AAB73550). The metalloproteases can be used for the treatment of
 XX cancers and arthritis. The invention also relates to the genes encoding
 XX MDT54 and MDT55, vectors and host cells containing the MDT54 or MDT55
 XX genes, the recombinant production of MDT54 and MDT55, and antibody
 XX specific for MDT54 or MDT55, and methods of screening for compounds
 XX which modulate the activity of MDT54 and/or MDT55. The present
 XX sequence represents a human ADAM-type metalloprotease-related DNA
 XX sequence, used in an exemplification as a probe for screening human cDNA
 XX libraries in order to isolate MDT54 cDNA.
 XX SQ Sequence 2670 BP; 511 A; 853 C; 867 G; 439 T; 0 other;
 Query Match 23.3%; Score 666; DB 22; Length 2670;
 Best Local Similarity 59.2%; Pred. No. 7.2e-129;
 Matches 1355; Conservative 0; Mismatches 885; Indels 48; Gaps 11;
 QY 157 CTCATTTTTCAGATCAGACATTTTCAGGAGACTTTTACCTACACCTGACGCGCGATGCT 216
 Db 157 CTCGCGCTCCACCTGCTCCGCTTCGGCAAGGGCTTCGTGCTGCGCTCGCGCGCGAGC 216
 QY 217 CAGTCTTTGGCTCCGCTTCTCCACTGAGCATCTGGGCTCCCGCTCCAGGGGCTCACC 276
 Db 217 AGCTTCTTGGGCGCGGAGTTCAAGATCGAGCGCTTCGGGGGCTCCGGCGGGCGC---ACC 273
 QY 277 GGGGCTCTTCAGACCTCGACGCTGCTTCTATTCTGGGGACGTGAACGCCGAGCCGAGC 336
 Db 274 GGGGCGAGCGGGGCTGCGCGGCTGCTTCTTCTCCGSCACCGCTGAATGGGAGCCCGAG 333
 QY 337 TCGTTTCTGCTGTGAGCTTGTGGGGGGGCTCGCGGAGGCTTTGGCTACCGAGGCGCC 396
 Db 334 TCGCTGGCGGGGCTGAGCTGTGCGGGGCTGAGCGGCTCTTCTTCTGCTGGAGCGGCGAG 393
 QY 397 GAGTATGTCATTAGCGCGCTGCCAAATGCTAGCGCGCGGGCGGCGAGCAACAGAGCAG 456
 Db 394 GAGTTTCACTCCAGCCCGAGGGCGCGGGGGCTCCCTGCTGAGCGCACCGGCTCGAG 453
 QY 457 GCGGCACACCTTCTCAGCGCGGGGGTGTTCGGGGCGGGGCTTCCGGAGAGCCACCTCT 516
 Db 454 CGCTGGGCTCCGCGGAGCGCCCGCTCCCGGAGGACCCGAGTGGGAGTGGAGAGC 513
 QY 517 CGCTCGGGGGTG- GCCTCGGGCTGGAACCCCGGCATCTACGGGCGCTTGGACCTTACAA 575
 Db 514 GGAGAGGGTTCAGAGCGAG 573
 QY 576 GCGCGCGCGGGGCTTTCGGGGAGAGTCTAGCGCGGGCAGG-----TCTGGCGGCGCC 630
 Db 574 GAAGAGAGGAGAGAGAGGCGCTAGCGAGCGCCGAGCCCGCTTGGGGGCGCACAGTAGAGCC 633
 QY 631 AAGCCGTTTCTGCTCTATCCCGGCTTACGTGGAGAGCTTGGTGGCGGAGCAGTCAATG 690
 Db 634 AAGCGGTTTGTCTCTGAGCGCGCTTCTGTAAGAGCTGCTGGTGGCGGAGTCTCCATG 693
 QY 691 GTCAAGTTCAGCGCGGAGCTGGAAACATTTATCTGTGAGCTGCTGGCAACCGCGGCG 750
 Db 694 GCTGCTTCTACGGGCGGAGCTGCAGAACACACATCTCTGAGCTTAATGCTGTGGAGCC 753
 QY 751 CGACTCTACCGCCATCCAGCATCTCAACCCCATCATCGTTGTGTCAAGTGTCTG 810

Db	754	CGAATCTACAAGCACCCCGCATCAAGAATTCCATCAACCTGATGGTGGTAAAGTGCCTG	813
Qy	811	CTTCTTAGAGATCGTGACTCCGGGGCCCAAGGTCAACGGCAATCGGGCCCTCAGCTCGCG	870
Db	814	ATCGTAGAAGATGAAATAATGGGGCCAGAGGTGTCCGACATATGGGGGCTTACATCGCT	873
Qy	871	AACTTCTGTGCGTGGCAGAGAAGTGAACAAGTGAAGTGAACAAGCAACCCGAGTACTGG	930
Db	874	AACTTCTGCAACTTGGCAGGGGGGTTTCAACCAAGCCAGCCGACCCAGAGCACTAC	933
Qy	931	GACACTGCCATCTCTTCAACGAGGAGGACCTGTGTGG---AGCCACCACTGTGACACC	987
Db	934	GACAGGCCATCTCTGTCAACAGACAGAACTTCTGTGGGAGGAGGGCTGTGTGACACC	993
Qy	988	CTGGGCATGTGCTGATGTGGGTACCATGTGTGACCCCAAGAAAGTGTCTGTTCATTGAG	1047
Db	994	CTGGGTGCGCAGACATCGGGACCATTTGTGACCCCAACAAAGAGTGTCTCGTGATCGAG	1053
Qy	1048	GACATGGGCTTCCATAGCCTTCACACTGCCACAGCTGGGCCAGCTGGCCACGTGTTCACATG	1107
Db	1054	GATGAGGGCTCCAGCGGCCACACCCCTGGGCCATGAATAGGCGACGCTCTCAGCATG	1113
Qy	1108	CCCATGACAACTGAAATCTGTGAGGAGTGTTTGGGAAGCTTCCGAGCAACCCACATG	1167
Db	1114	CCCAAGCAGACTCCAAGCCCTGCACAGGCTTCTTCGGGGCCCATGGGCAAGCAACGCTG	1173
Qy	1168	ATGTCCCGGACCCCTCATCCAGATCGACCGTGCACACCCCTGGTCAGGCTGCAGTGTGCC	1227
Db	1174	ATGCAACCGCTGTTCTGTCACCTGAACAGAGCGTGCCTGTCTCCCTCGAGGCGCATG	1233
Qy	1228	ATCATCAGCACTTCTTGACAGCGGGCACGGTGAAGTGCCTTCCTGGACCAACCCAGCAAG	1287
Db	1234	TATCTCACAGAGCTTGTGACGGCGGGCACGAGACTGTCTCTGGATGCCCTGTCTGCG	1293
Qy	1288	CCCATCTCCCTGCCGAGGATCTGCCGGG-----GCCAGCTACACCTTGAGCCAGCAG	1341
Db	1294	GCCCTGCCCTTCCCAACAGGCTTCCGGGCGCATGGGCCCTGTACCAAGCTGGACCAAGCAG	1353
Qy	1342	TGCGAGTGGCTTTTGGCGTGGGCTCCAAGCCCTGCTCTTACA-----TGCAGTAC	1392
Db	1354	TGCAGGCAGATCTTTGGGCGGATTTCCGCCACTGCCCAACACCTCTGCTCAGGACGTC	1413
Qy	1393	TGCACCAAGCTGTGGTG---CACCGGGAAGGCCAAGGACAGATGGTGTGCCAGACCCGC	1449
Db	1414	TGGCGCCAGCTTTGGTGTCCACACTGATGGGCTGAGCCCTGTGCCACACGAAGATGGC	1473
Qy	1450	CACCTCCCTGGGCGATGGCACCAAGCTGTGGCAGGGCAAGCTCTCCCTCAAGGGGCC	1509
Db	1474	AGCCTGCCCTGGGCTGACGGCACGCGTGTGGGGCGCTGGCAACCTCTGCTCAGAAGGCAGC	1533
Qy	1510	TG-----CGTGGAGAGACACAACCTCAACAAAGCACAGGTTGGATGTTCTCTGGGCCAA	1563
Db	1534	TGCTACCTGAGGAGGAAGTGAGAGGCCCCAAGCCCTGTAGATGAGGCTGGGCACCG	1593
Qy	1564	TGGGATCCCTATGGCCCCCTGCTCGCGACATGTGTGGGGCGCTGCAGCTGGCCAGGAGG	1623
Db	1594	TGGGGACCTCTGGGAGAAATGTTCTCGACCTGTGGAGGAGGATACAGTTTTTACACCCGT	1653
Qy	1624	CAGTGCACCAACCCACCCCTGCCAACGGGGCAAGTACTGCGAGGAGGTGAGGGTGA	1683
Db	1654	GAGTGCAGGACCCGAGGCTTCAGAAATGGAGGAAGATACTGCCTGGGTGGAGAGCAAG	1713
Qy	1684	TACCGATCTGCAACTGGAGCCCTGCCCAAGCTCAGCCTCCGGAAGAGCTTCCGGGAG	1743
Db	1714	TACCACTATGCCACAGGAGGAATGCC-----CCTGACGGGAAGAGCTTCAGGGAG	1767
Qy	1744	GAGCACTGTGAGGCTTTCAACGGCTTACAACACAGCAACCAACCGGCTCACTCTCGCGCTG	1803
Db	1768	CAGCACTGTGGAAGTATAATGSCCTTACAATTTACACTGACATGGACGGGAATCT---CTTG	1824
Qy	1804	GCATGGGTGCCAAGTACTTCGGCGTGTCTCCCGGGAAGTGAAGCTCATCTGCCGA	1863

Db	1825	CAGTGGGTCCCAAGTAATGCTGGGGTGTCC	CCCCCGGACCGCTGCAAGTTGTTCTTCCGGA	1884
QY	1864	GCCAAATGGCACTGCTACTTCTATGTCTGGC	ACCCCAAGSTGTGGACGGCAGCTGTGC	1923
Db	1885	GCCCGGGGAGGACCGAGTTCAAGATGTTCTG	AGGCCCAAGGTGATTGATGGTACCTCTGTGT	1944
QY	1924	TCTCCTGACTCCACCTCCGCTCTGTGTCCAA	GGAAGTGCATCAAGGCTGCTGTGATGGG	1983
Db	1945	GGCCAGAAACACTGGCCACTCTGTGTCCGTG	GGCCAGTGTGTCAAGGCCGGTGTGACCAT	2004
QY	1984	AACCTGGGCTCCAAAGAAGAGATTCTGCACA	AGTGTGGGGTGTGTGGGGGAGACAAATAAG	2043
Db	2005	GTGTGTGACTCGCCTCGGAAGCTGGACAAAT	TCGGGGTGTGTGGGGGCAAAAGCAACTCC	2064
QY	2044	TGCAAGAAGGTGACTGGACTCTTTCACCAAG	CCCATCATGCGCTACAATTTCTGTGTGGCC	2103
Db	2065	TGCAGGAAGGTCTTCGGGTCCCTTCA	CCCCACCAATATTGGGTACAATGACATTTGT	2124
QY	2104	ATCCCCGAGCGGCTCAAGCATCGACATCCG	CCAGCGGTTTACAAAGGGCTGATCGG	2163
Db	2125	ATCCACGCTGGTGCCTAATATTGACGTCA	AGCAGCGGAGCCACCCGGGTGTGCAGAAC	2184
QY	2164	GATGACAACCTACCTGGCTCTGAAGAACAC	CGCCAAAGCAAGTACTGCTCAAGCGGCATTTC	2223
Db	2185	GATGGAACTACCTGGCGCTGAAAGACGGCT	GATGGGCACTGCTCTCAAGGGCAACCTG	2244
QY	2224	GTGTGTGCGGCTGGAGCGGGACCTGTGTGT	GTAAGGCGAGTCTGCTCGGTACAGCGGC	2283
Db	2245	GCCATCTCTGCCATAGACAGGACATCTTGT	GTAAGGGAGCCATCCTGAAGTACAGCGGC	2304
QY	2284	ACGGGCACAGCGGTGGAGAGCCTTGCAGG	CTTCCCGGCCCATCCTGTAGGCGCTGACCGTG	2343
Db	2305	TCCATGCCACCTCGGAGCGCTTGCAGAGCT	TCCGGCCCTTGGCAGAGCCTCTGACAGTG	2364
QY	2344	GAGTCTCTTCCTGTT--GGGAGATGACAC	CGCCCGGGTCCGCTACTCTCTTCTATCTG	2400
Db	2365	CAGTCTCTGACGTCCCTGGCGAGTCTTCC	CCCCCAAAAGTCAAAATACACCTTCTTTGTT	2424
QY	2401	CCCAAGA 2408		
Db	2425	CCTAATGA 2432		
RESULT	11			
AAAF82166				
ID	AAAF82166	standard; cdna; 2670 BP.		
XX				
AC	AAAF82166;			
XX				
DT	02-JUL-2001	(first entry)		
XX				
DE	Human ADAM type metal protease MDP53	encoding cdna SEQ ID NO:21.		
XX				
KW	Human; a disintegrin and metalloprotease	type metal protease; MDP51;		
KW	MDP52; MDP53; ADAM type metal protease;	cytostatic; antiarthritis;		
KW	cancer; arthritis; arthrosis deformans;	ss.		
XX				
OS	Homo sapiens.			
XX				
FH	Key	Location/Qualifiers		
FT	CDS	1..2670		
FT		/*tag= a		
FT		/product= "MDP53"		
FT		/note= "a disintegrin and metalloprotease (ADAM)		
FT		type metal protease"		
XX				
XX	JP2001008687-A.			
XX				
PD	16-JAN-2001.			
XX				
PF	25-JUN-1999;	99JP-0180973.		
XX				
PR	25-JUN-1999;	99JP-0180973.		

QY 1984 AACCTGGCTCAAGAAGAGATTGCAACAAGTGTGGGGTGTGGGGAGACAATAAGACG 2043
 DB 2005 GTGGTGGACTCGCTCGAAGCTGGACAATGCGGGGTGTGGGGCAAGGCAACTCC 2064
 QY 2044 TGAAGAAGGTGACTGGAGCTCTTACCAAGCCCATGATGGCTACAAATTTCTGGTGGCC 2103
 DB 2065 TGCAGGAAGGTCTCGGGTCCCTCACCCCAACCAATATGGCTACAAATGACATTTGTCACC 2124
 QY 2104 ATCCCCCAGCGGCTCAAGCATCGCATCCGCGGCTTACAAAGGGCTGATCGGG 2163
 DB 2125 ATCCCACTGTGCGCACTAATATTGAGTGAAGCAGCGGACCCCGGGTGTGCAGAAC 2184
 QY 2164 GATGACAACCTACCTGGCTGTGAAGACAGCCAGGCAAGTACCTGTCAACGGGCATTTC 2223
 DB 2185 GATGGAACTACCTGGCGCTGAAGCGGCTGATGGCAGTACTGTCTCAAGGCAACCTG 2244
 QY 2244 GTGGTGTGCGGGTGGAGCGGACCTGGTGGTGAAGGGCAGTGTCTGGGTACACGGCG 2283
 DB 2245 GGCATCTCTCCATAGACAGGACATCTTGGTGAAGGGGACCATCTCTGAAGTACAGCGCG 2304
 QY 2284 ACGGGCACAGCGGTGGAGACCTGCAAGCTTCCCGGCCATCTCTGGAGCGCTGACCGNG 2343
 DB 2305 TCCATCGCCACCTGGAGCGCTCGAGAGCTTCCGGCCCTTGCAGAGCTCTGACAGTG 2364
 QY 2344 GAGTCTCTCTCGT---GGGGAAGATGACACCGCCCGGGTCCGCTACTCTTCTATCTG 2400
 DB 2365 CAGCTCTGACAGTCCCTGGCGAGGTCTTCCCCCAAAAGTCAATACACCTTCTTTGTT 2424
 QY 2401 CCCAAAGA 2408
 DB 2425 CCTAATGA 2432

RESULT 12

AAZ32001
 ID AAZ32001 standard; cdna; 3008 BP.

XX AC AAZ32001;

XX DT 10-JAN-2000 (first entry)

XX DE Human METH2 encoding cdna.

XX KW Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
 cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
 angiogenesis inhibitor; abnormal wound healing; inflammation;
 rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
 diabetic retinopathy; macula degeneration; haemangioma; detection;
 arterial-venous malformation; immune deficiency; ss.

XX OS Homo sapiens.

XX PN WO937660-A1.

XX PD 29-JUL-1999.

XX PF 22-JAN-1999; 99WO-US01313.

XX PR 23-JAN-1998; 98US-0072298.

XX PR 28-AUG-1998; 98US-0098539.

XX PA (IRUE/) IRUELA-ARISPE L.

XX PA (HAST/) HASTINGS G A.

XX PA (RUBE/) RUBEN S M.

PI IrueLa-Arispe L, Hastings GA, Ruben SM;

XX WPI; 1999-590684/50.

DR P-PSDB; AAY49502.

XX New isolated metalloprotease thrombospondin polypeptides, useful for

PT treating hyperproliferative disorders, cancers or autoimmune disorders

XX PS

Claim 4; Fig 2; 457bp; English.

XX CC

AAZ32000 and AAZ32001 encode, and AAY49501 and AAY49502 represent, human metalloprotease thrombospondin (METH) proteins METH1 and METH2 respectively. METH1 and METH2 have been found to be potent inhibitors of angiogenesis both in vitro and in vivo. They can be used for treating cancer and other disorders related to angiogenesis including abnormal wound healing, inflammation, rheumatoid arthritis, psoriasis, endometrial bleeding disorders, diabetic retinopathy, some forms of macula degeneration, haemangiomas, and arterial-venous malformations. They may be useful in treating deficiencies or disorders of the immune system by activating or inhibiting the proliferation, differentiation, or mobilisation (chemotaxis) of immune cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or toxins), or infectious. They can also be used to treat inflammatory conditions, both chronic and acute conditions. The products can also be used for detection and diagnosis. AAZ32002 to AAZ32080, and AAY49503 to AAY49511 represent sequences given in the exemplification of the present invention.

XX SQ Sequence 3008 BP; 617 A; 924 C; 948 G; 514 T; 5 other;

Query Match 23.3%; Score 666; DB 20; Length 3008;

Best Local Similarity 59.2%; Pred. No. 7.3e-129;

Matches 1355; Conservative 0; Mismatches 885; Indels 48; Gaps 11;

QY 157 CTCATTTTTCAGATCACAGCATTTTCAGGAGGACTTTTACCTACACCTGATGCCCGATGCT 216

DB 160 CTCGGCTCCACCTGTCCGCTTCGGCAAGGCTTCGTGTGGCTGGCGCCGACGAC 219

QY 217 CAGTTCCTGGTCCGCTTCTCCACTGAGCATCTGGGCTCCCTCCAGGGGTCAAC 276

DB 220 AGCTTCCTGGGCGCGGCTTCAAGATCGAGCGCTTCGGGCTCCGGCGGCG---ACC 276

QY 277 GGGGCTCTTCAGACCTCGAGCGCTGTCTTATTCGGGACGTGAAGCCGAGCCGAC 336

DB 277 GGGGCGAGCGGGCTCGCGGCTGTCTTATTCGGGACGTGAAGCCGAGCCGAC 336

QY 337 TCGTTCGTGTGTAGCCTGTGGGGGGCTCCCGGAGCCTTTTGGCTACCGAGGCGCC 396

DB 337 TCGTTCGTGTGTAGCCTGTGGGGGGCTCCCGGAGCCTTTTGGCTACCGAGGCGCC 396

QY 397 GAGTATGTCATTAGCCCTCCCAATCTAGCGCGCGCGCGCGAGCGCAACAGCCAG 456

DB 397 GAGTATGTCATTAGCCCTCCCAATCTAGCGCGCGCGCGCGAGCGCAACAGCCAG 456

QY 457 GCGCACACCTTCTCCAGCGCGGGGTGTTCGGGCGGGCTTCGGAGAGACCCACCTCT 516

DB 457 GCGTGGGTCCCGCGGAGCGCCCTCCCGGAGACCCGAGTGGAGGTGGAGACG 516

QY 517 CGCTCGGGGTG-GCCTCGGGCTGGAACCCCGCCATCTACGGGCCCTTGACCCCTTCAA 575

DB 517 GGAGAGGTTCAGAGGCGAGGAGAGAGAGACCAACAGGAGGACAGGAGGAGGACCAA 576

QY 576 GCGCGCGGGCGGGCTTCGGGGGAGAGTCTGTAGCGCGCAGG-----TCTGGCGGCC 630

DB 577 GAAGAGGCGAGAGAGGCGCTAGGAGCGCCGCCCGCCCTGGGGCGCAGAGTAGAGACC 636

QY 631 AAGCGTTTGTGTATATCCCGCGGTACGTGGAGAGCGTGTGTGTGGCGGACGATCAATG 690

DB 637 AAGCGTTTGTGTCTGTAGGCGCGCTTCTGTGGAGAGCGTGTGTGTGGCGGATGCGCTCCATG 696

QY 691 GTCAAGTTCCAGCGCGGACCTGGAACATTTCTGTCTGAGCTGTCTGGCAACGGCGGG 750

DB 697 GCTGCTTCTACGGGGCGGACCTGCAGAACACATCTCTGACGTAAATGTCTGTGCGACGC 756

QY 751 CGACTCTACCGCCATCCAGCATCTCAACCCCAACATCGTTGTGTCAAGGTGCTG 810

DB 757 CGATCTCAAGACACCCCGACATCAAGATTCATCAACCTGATGGTGGTAAATGTGCTG 816

QY 811 CTTCTTAGAGATGTGACTCCGGGCCCAAGGTACACCGCAATGCGGCCCTGACGCTGCGC 870

Db 817 ATCGTAGAAGTGAATAATGGGCCCCAGAGGTGTCCGCAATAGTGGGGGCTTACACTGCGT 876
Qy 871 AACTTCTGTGCTGGGAGAGAGCTGACACAAAGTGAAGTACACAGACACCCGAGTACTGG 930
Db 877 AACTTCTGCAACTGGGAGCGCGTTTCAACACCCAGGAGCGCCACCCAGAGCACTAC 936
Qy 931 GACACTGCCATCTCTTACACAGGAGGACCTGTGTGG---AGCCACCACCTGTGTACACC 987
Db 937 GACAGGCCATCTCTGTCACACAGACAACTTCTGTGGCAGGAGGGGTGTGTACACC 996
Qy 988 CTGGCATGCTGATGTGGGTACCATGTGTGACCCCAAGAGAGCTGCTGTGTCATTGAG 1047
Db 997 CTGGTGTGCAGACATCGGAGCAATTTGTGACCCCAACAAAGCTGCTCCGTGTGTCAG 1056
Qy 1048 GAGATGGGTCTCATCAGCCTTCAACAGCTGACAGTGGCCAGCTGGCCAGCTGTTCACATG 1107
Db 1057 GATGAGGGGTCCAGGGGGCCACACCTTGGCCCATGAACTAGGGCACCTCTCTCAGCATG 1116
Qy 1108 CCCCATCACAATGTGAAGTCTGTGAGGAGGTGTGGAAGCTCCGAGGCAACACCATG 1167
Db 1117 CCCACAGACATCCAAAGCCTGCACAGCTCTTGGGGCCATGGGCAAGCACACGTG 1176
Qy 1168 ATGTCCCGACCTTATCCAGATCGAACCTGTCACCAACCCCTGTGACGCTGACGTGGCC 1227
Db 1177 ATGGCACCGCTTTCGTGTCACCTGAACACAGACGCTGCCCTGTGTCCCTCTGCAGCGCATG 1236
Qy 1228 ATCATCACCACTCTCTGGACAGGGGACGCTGACTGCTCTCTGACCAACCCACCAAG 1287
Db 1237 TATCTCAGAGCTTCTGGACGGGGGACGAGACTGTCTCTGATGGCCCTTGTGTGG 1296
Qy 1288 CCCATCTCCCTGCCCGAGGATCTGCCGGG-----GCCAGCTACACCTCTGACCCAGCAG 1341
Db 1297 GCCCTGCCCTCCACAGGCTCCCGGGCGGATGSCCTGTACCAGCTGGACACGAGCAG 1356
Qy 1342 TGCAGCTGGCTTTGGCGTGGGTCTCAAGCCCTGTCTTACA-----TGCAGTAC 1392
Db 1357 TGCAGGAGAGATCTTTGGCGCGGATTTCCGCCACTGTCCGCCAACACACCTGTGCTCAGGACGTC 1416
Qy 1393 TGCACCAAGCTGTGGTG---CACCGGAAGGCAAGGACAGATGTTGCCAGACCCCGC 1449
Db 1417 TGCGCCAGCTTTGGTCCACACTGTGGGCTGAGGCTGAGCCCTGTGCCACAGAAATGCG 1476
Qy 1450 CACTTCCCTGGCCGATGGACACAGCTGTGGGAGGGGCAAGCTCTGCCCTCAAGGGGCC 1509
Db 1477 AGCCTGCCCTGGCTGACGGCAGCCCTGGGGCCTGGGACCTCTGCTCAGAGGACGAGC 1536
Qy 1510 TG-----CGTGGAGAGACAACTCAACAAGCACAGGTTGGTGTCTTGGGCCAAA 1563
Db 1537 TGTCTACTGAGGAGGAAGTGGAGGCCCAAGCCCTGTGTAGATGGAGCTGGCCACCG 1596
Qy 1564 TGGGATCCCTATGGCCCTGTCTCGGCACATGTGGTGGGGCGCTGACGCTGGCCAGGAGG 1623
Db 1597 TGGGACCTCTGGGAGATGTTCTCGACCTGTGGAGGAGGTACAGTTTTCACACCGT 1656
Qy 1624 CAGTGCACCAACCCCTCTGCCAACGGGGCAAGTACTGCCAGGAGTACAGGTGAGAA 1683
Db 1657 GAGTGCAGAGGACCCGAGCCTCAGAATGGAGGAATATCTGCTGGTGGGAGGACCAAG 1716
Qy 1684 TACCATCTCCGACCTGGAGCCCTCGCCAGCTCAGCCTCCGGAAAGAGCTTCCGGGAG 1743
Db 1717 TACCAGTATGCCACAGGAGGAATGCC-----CCTGACGGGAAAGACTTCAGGGAG 1770
Qy 1744 GAGCAGTGTGAGGCTTTCAACGGCTACAAACACAGCACACCGGCTCACTCTCGCGGTG 1803
Db 1771 CAGCAGTGTGAGAAGTATAATGCTACAAATTTACACTGACATGGACGGGAATCT---CCTG 1827
Qy 1804 GCATGGGTGCCCAAGTACTCTGGCGGTGTCTCCCGGGACAAAGTCAAGCTCACTGCGCGA 1863
Db 1828 CAGTGGGTCCCCAAGATGTCTGGGGTGTCCCGGGGACCGCTCAAGTGTCTGCGCGA 1887
Qy 1864 GCCAATGCACTGGCTACTTCTATGTCTGCGACCCCAAGTGTGTGGACGGCAGCCTGTC 1923

Db 1888 GCCCGGGGAGGAGGAGAGTTCAAAGTGTTCAGGCCCAGGTGATGATGGGACCTGTGT 1947
Qy 1924 TCTCTGTACTCCACCTCGCTGTGTCCAGGCAAGTGCATCAAGGCTGGTGTGTGATGG 1983
Db 1948 GGGCAGAAACACATGGCCATCTGTGCGTGGCAGTGTGTCAAGCGCGCTGTGACCAT 2007
Qy 1984 AACCTGGCTCCAAGAAGATTCGACAAGTGTGGGTGTGTGGGGAGACAATTAAGAGC 2043
Db 2008 GTGTGGTACTCGCTCGGAAGCTGGACAATCGCGGTGTGTGGGGGCAFAAGGCAACTCC 2067
Qy 2044 TCCAGAAGGTGACTGGACTCTTACCAAGCCCATGATGCTCAATTTCTGCTGTGGCC 2103
Db 2068 TGCAGGAAGGTCTCGGCTCCCTCACCCACCAATTAATGCTCAATGACATTTGTAC 2127
Qy 2104 ATCCCGCAGGGCTCAAGCATCGACATCCGCGCAGCGGTTTACAAAGGCTGTATCGG 2163
Db 2128 ATCCAGCTGTGGCCTCACTAATATTGACCTGAAGCAGCGGAGCCACCGGGTGTGAGA 2187
Qy 2164 GATGACAATACCTGGCTCTGAAGAACAGCAAGTACCTGCTCAACGGGCAATTC 2223
Db 2188 GATGGAACTACTGGCTGAAGACGCTGATGGGCACTACCTGCTCAACGGCAACCTG 2247
Qy 2224 GTGGTCTCGGCGTGAAGGACCTGTGTGAAGGCGAGTCTGCTGCGGTACAGCGGC 2283
Db 2248 GCATCTCTGCCATAGACGAGCATCTTGTGAAGGGGACCATCTCTGAAGTACAGCGGC 2307
Qy 2284 AGGGCAGACGGGTGGAGAGCTCGAGGCTTCCCGGCCCATCTCTGGAGCCGCTGACCGTG 2343
Db 2308 TCCATCGCACCTCGAGCGCTGCAGAGCTTCGCGCCCTTGCCAGAGCCTCTGACAGTG 2367
Qy 2344 GAGGTCTCTCGT---GGGGAAGATGACACCGCCCGGGTCCGCTACTCTCTATCTG 2400
Db 2368 CAGCTCTCTGACAGTCCCTGGGAGGCTTCCCGCCCAAAAGTCAATACACCTTCTTGT 2427
Qy 2401 CCCAAGA 2408
Db 2428 CCTAATGA 2435
RESULT 13
AAC90058
ID AAC90058 standard; DNA; 3008 BP.
XX AAC90058;
AC AAC90058;
XX 19-MAR-2001 (first entry)
XX Human METH2 coding sequence.
XX Human; METH2; metalloprotease; thrombospondin; angiogenesis inhibition;
KW cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischaemic limb angiogenesis; Osler-Webber syndrome;
KW plaque neovascularisation; telangiectasia; haemophilic joint;
KW angiofibroma; fibromuscular dysplasia; wound granulation;
KW Crohn's disease; atherosclerosis; birth control; ss.
XX Homo sapiens.
OS WO200071577-A1.
XX 30-NOV-2000.
XX 25-MAY-2000; 2000WO-US14462.
XX 25-MAY-1999; 99US-0318208.
PR 20-JUL-1999; 99US-0144882.
PR 10-AUG-1999; 99US-0147823.
PR 13-AUG-1999; 99US-0373658.
PR 22-DEC-1999; 99US-0171503.

22-FEB-2000; 2000US-0183792.
 (HUMA-) HUMAN GENOME SCI INC.
 (SMIK) SMITHKLINE BEECHAM CORP.
 (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 (IRUE/) IRUELA-ARISPE L.
 (HAST/) HASTINGS G A.
 (RUBE/) RUBEN S M.
 (JONA/) JONAK Z L.
 (TRUL/) TRULLI S H.
 (FORN/) FORNWALD J A.
 (TERR/) TERRETT J A.

IrueLA-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
 Fornwald JA, Terrett JA;
 WPI: 2001-025136/03.
 N-PSDB; AAC90058.

METH1 and METH2 polynucleotides and encoded polypeptides, used to inhibit angiogenesis in the treatment of disorders such as cancer, rheumatoid arthritis and psoriasis -
 Claim 14; Fig 2; 768pp; English.

The present sequence is the coding sequence of human METH2 (ME for metalloproteinase and TH for thrombospondin). METH2 can be used for inhibiting angiogenesis in an individual, and for treating cancer, benign tumours, an ocular angiogenic disease, rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis, vasculogenesis, granulations, hypertrophic scars, nonunion fractures, scleroderma, trachoma, vascular adhesions, myocardial angiogenesis, coronary collaterals, cerebral collaterals, arteriovenous malformations, ischaemic limb angiogenesis, Osler-Webber syndrome, plaque neovascularisation, telangiectasia, haemophilic joints, angiofibroma, fibromuscular dysplasia, wound granulation, Crohn's disease or atherosclerosis. METH2 can also be used in birth control. METH2 can also be used in diagnostic methods for the prognosis of cancer.

Sequence 3008 BP; 617 A; 924 C; 948 G; 514 T; 5 other;

Query Match 23.3%; Score 666; DB 22; Length 3008;
 Best Local Similarity 59.2%; Pred. No. 7.3e-129;
 Matches 1355; Conservative 0; Mismatches 885; Indels 48; Gaps 11;

QY 157 CTCATTTTTCAGATCACAGCATTTTCAGGAGGACTTTTACCTACACCTGACGCGCGATGCT 216
 DB 160 CTCGCGCTCCACCTGTCGCGCTTCGCAAGGCTTCGTGTTGCGCTTGGCGCCGACGAC 219
 QY 217 CAGTTCTTGGCTCCCGCTTCCACTGAGCATCTGGGGTCCCGCTCCAGGGGCTCACC 276
 DB 220 AGCTTCTGGCGCGGAGTCAAGATCGAGCGCTTCGGGGGCTCGGGCGCGGCG---ACC 276
 QY 277 GGGGGCTTTCAGACTGCGAGCGTCTTCTATTCTGGGAGCTGAACCGCGAGCCGAC 336
 DB 277 GGGGGGAGCGGGCTGCGCGCTGTTTTTTCGGCAGCGTGAATGGGAGCCCGAG 336
 QY 337 TCGTTGCTGTGTAGCGCTGCGGGGGCTTCGCGAGCGCTTGGCTACGAGCGGCC 396
 DB 337 TCGCTGGCGGGGTGAGCTGTGCGCGGGCTGAGCGGCTCTTCTGCTGGACGGCGAG 396
 QY 397 GAGTATGTCATTAGCCCGTCCCAATGCTACGCGCGCGGGCGCGAGCGCAACAGCCAG 456
 DB 397 GAGTTACCATCCAGCGCGAGGCGCGGGGGCTCCCTGGCTCAGCGCGACCGCTCGAG 456
 QY 457 GCGCACACCTTCTCCAGCGCGGGGTGTTCCGGGGGGGCTTCGCGAGAGCCCACTCT 516
 DB 457 CGCTGGGTTCGCGGAGCCCGCCCTCCCGGAGGAGCCGAGTGGGAGTGGAGAGC 516
 QY 517 CGCTCGGGGTG-GCCTCGGGTGAACCCCGCCCATCTCAGGGCGCTTGAACCTTACAA 575
 DB 517 GGAGAGGCTCAGAGCAGGAGAGAGAGACACACAGGAGGAGGAGGAGGAGGAGGAGCA 576

QY 576 GCCGGCGGGCGGCGCTTCGGGGAGAGTCTGAGCGGGCGCAGG-----TCTGGCGGGGCC 630
 DB 577 GAAGAGGAGGAGAGAGGCGCTAGCGAGCGCCACCGCCCTGGGGCCCAVAGTAGGACC 636
 QY 631 AAGCGTTTCTGCTCTATCCCGGCTACCTGGAGAGCGTGGTTCGCGGAGCAGG*CAATG 690
 DB 637 AAGCGTTTCTGCTCTGAGAGCGGCTTCGTGGAGAGCGTCTGGTGGCGGATGCGCTTCCATG 696
 QY 691 GTCAAGTTCCACGGCGCGGACCTGGAACTTATCTGTGACGCTGTCTGGCAACGGCGCG 750
 DB 697 GCTGCTTCTACGGGGCGGAGCTGCAGAAACACACTCTCTGACGTTAATCTCTGTGGCAGCC 756
 QY 751 CGACTCTACCGCCATCCAGCATCTCAACCCCATCAACATCTGTGGTCAAGTGGTGT 810
 DB 757 CGAATCTAAGACACCCCGCATCAAGAAATTCATCAACCTGATGGTGGTAAAAAGTGTG 816
 QY 811 CTTCTTAGAGATCTGACTCCGGGCGCAAGGTCAACCGGCAATCGCGCCCTCAACCTCGCG 870
 DB 817 ATCGTAGAAGATGAAAAATGGGCCAGAGGTGTCCGACATGGGGCTTACACTCGGT 876
 QY 871 AACTTCTGTGCTGGCAGAAAGCTGAACAAAGTGAAGTGAACAGCACCCTGAGTACTGG 930
 DB 877 AACTTCTGCAACTGGCAGCGCGTTTCAACCGAGCGCCAGCGCCACCCAGAGCACTAC 936
 QY 931 GACACTGCATCTCTTCAACAGCAGGAGCTGTGTGG---AGCCACACCTGTGACACC 987
 DB 937 GACAGCGCATCTCTGCTCAACAGACAGACTCTGTGGGAGGAGGGCTGTGTGACACC 996
 QY 988 CTGGCATGGCTGATGTGGGTACCATGTGTGACCCCAAGAGAAAGCTCTCTGTCTATTGAG 1047
 DB 997 CTGGGTGTGGCAGACATCGGACCATTTGTACCCCAACAAAGCTCTCTGCTGATCGAG 1056
 QY 1048 GACATGGCTTCCATCAGCTTTCACACCTGCCCCAGAGCTGGGCCAGCTGTTCACATG 1107
 DB 1057 GATGAGGGCTTCCAGGCGGCCACACCTTGCCCATGAATAGGAGCACTCTCAGAGT 1116
 QY 1108 CCCATGACAAATGTAAAGTCTGTGAGGAGGTGTGGGAAAGCTTCCGAGCAACCAATG 1167
 DB 1117 CCCACGAGCACTCAAGCCCTGCACAGCGCTCTTCGGGCCCATGGGCAAGCACCAGTG 1176
 QY 1168 ATGTCCCGGACCTTCATCCAGATCGCGTGGCCAAACCCCTGGTGAGCTGAGTGTGCC 1227
 DB 1177 ATGCACCGCTGTTCGTCCACTGAACAGCAGACGCTGCTCCCTGCTGAGCGCCATG 1236
 QY 1228 ATCATCACCGACTTCTCTGAGCGGCGGCGGAGAGAGTGTCTCTGGATGCCCTGTGGTGG 1287
 DB 1237 TATCTACAGAGCTTCTGGAGCGGCGGCGGAGAGAGTGTCTCTGGATGCCCTGTGGTGG 1296
 QY 1288 CCCATCTCCCTGCGGAGGATCTGCCGGGC-----GCCAGCTACACCTGAGCCAGCAG 1341
 DB 1297 GCGCTGCCCTCCCAACAGGCTCCCGGGCGCATGGCCCTGTACCAGCTGGACAGCAG 1356
 QY 1342 TGCAGCTGGCTTTTGGGTGGGTCTCAAGCCCTGCTCTTACA-----TGCAGTAC 1392
 DB 1357 TGCAGGAGAGATCTTGGGCGCGGATTTCCGCCACTTCCGCCCAACACCTCTGCTGCTG 1416
 QY 1393 TGCACCAAGCTGTGTGTG---CACCGGAAGCCCAAGGAGCAGATGGTGTGCACACCCCG 1449
 DB 1417 TGCGCCAGCTTGTGTGCCACACTGATGGGGCTGAGGCCCTGTGCCACACCT*GAATGGC 1476
 QY 1450 CACTTCCCTGGGCGGATGGCACAGCTGTGGGAGGGCAAGCTCTGCCCTCAAGGGGCC 1509
 DB 1477 AGCTGCCCTGGGCTGACGCGCAGCCGCTGGGGCTTGGGCACTCTGCTCAGAAGGAGC 1536
 QY 1510 TG-----CGTGGAGACACACCTCAACAGCAGAGGTGGATGTTCTCTGGGCCAAA 1563
 DB 1537 TGTCTACTGAGGAGGAAGTGGAGAGCCCAAGCCCGTGTAGTAGAGCTGGGCAACC 1596
 QY 1564 TGGATCCCTATGCGCCCTGCTCGCGCACATGTGTGGGGGCTGCGAGCTGGCCAGGAG 1623
 DB 1597 TGGGAGCCCTGGGAGAGATGTTCTCGGACCTGTGGAGAGGAGTACAGTTTTCACACCGT 1656
 QY 1624 CAGTGCACCAACCCCAACCCCTTGCACAGGGGCAAGTACTCGAGGGAGTGGAGGTGAAA 1683

QY	2062	CTCTTCACCAAGCCCATGCATGCTACAAATTTTCGTGGTGGCCATCCCGCGA3GGCGCTCA	2122
Db	2580	ATAGTCTCTAGTACAAGACCTGGGTATCATGACATGTGCAAAATTTCTGTGGAGCCACC	2639
QY	2122	AGCATGACATCCGCCAGGCGGTTACAAAGGGCTGATCGGGGATGACAACTACCTGGCT	2181
Db	2640	AACATTGAAGTGAACATCGGAATCAAGGGGGTCCAGAAACAATGGCAGCTTTCTGGCT	2699
QY	2182	CTGAAGAACACCAAGCGAAGTACCTGCTCAAGGGCATTTCTGGTGTCTGGCGGTGGAG	2241
Db	2700	ATTAGAGCCGCTGATGGTACCTATATTTCTGAATGGAAACTTCACTGTCTCCACACTAGAG	2759
QY	2242	CGGCACTGTGTGTAAGGGCACTGCTGCGGTACAGCGGCACGGGCACAGCGGTGGAG	2301
Db	2760	CAAGACCTCACCTACAAAGGTACTGCTTTAAGGTACAGTGGTTCTCGGCTGGCCCTGAA	2819
QY	2302	AGCTTGAGGCTTCCCGGCCCATCTCTGGAGCGCTGACCGGTGGAGTCTCTCTCGGTGGGG	2361
Db	2820	AGATCCGCAAGCTTTTASTCCACTCAAGAAGAACCTTTAACCATCCAGGTTCTTATG3TAGGC	2879
QY	2362	AAGATGACACCGCCCGGGTCCGCTACTCTTCTATCTGCCCCAAGAGCCTCGGGAGGAC	2421
Db	2880	CATGCTCTCGACCCAAAATTAATTCACCTACTTTATGAAGAAGACAGAGTCAATTC	2939
QY	2422	AAGTCTCTTC 2431	
Db	2940	AACGCCATT 2949	
RESULT 15			
AAC90067			
ID	AAC90067 standard; DNA; 4180 BP.		
XX			
AC	AAC90067;		
XX			
DT	19-MAR-2001 (first entry)		
XX			
DE	D57076 cDNA clone.		
XX			
KW	METH; metalloprotease; thrombospondin; angiogenesis inhibition;		
KW	cancer therapy; benign tumour; ocular angiogenic disease;		
KW	rheumatoid arthritis; psoriasis; wound healing; endometriosis;		
KW	vasculogenesis; granulation; hypertrophic scar; nonunion fracture;		
KW	scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;		
KW	coronary collateral; cerebral collateral; arteriovenous malformation;		
KW	ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;		
KW	plaque neovascularisation; telangiectasia; haemophilic joint; EST;		
KW	angiofibroma; fibromuscular dysplasia; expressed sequence tag;		
KW	Crohn's disease; atherosclerosis; birth control; ss.		
XX			
OS	Unidentified.		
XX			
PN	WO200071577-A1.		
XX			
PD	30-NOV-2000.		
XX			
PF	25-MAY-2000; 2000WO-US14462.		
XX			
PR	25-MAY-1999; 99US-0318208.		
PR	20-JUL-1999; 99US-0144882.		
PR	10-AUG-1999; 99US-0147823.		
PR	13-AUG-1999; 99US-0373658.		
PR	22-DEC-1999; 99US-0171503.		
PR	22-FEB-2000; 2000US-0183792.		
XX			
(HUMA-) HUMAN GENOME SCI INC.			
PA	(SMK-) SMITHKLINE BEECHAM CORP.		
PA	(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.		
PA	(IRUE/) IRUELA-ARISPE L.		
PA	(HAST/) HASTINGS G A.		
PA	(RUBE/) RUBEN S M.		
PA	(JONA/) JONAK Z L.		
PA	(TRUL/) TRULLI S H.		

PA (FORM/) FORNWALD J A.
 PA (TERR/) TERRETT J A.
 XX Iruela-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
 PI Fornwald JA, Terrett JA;
 XX WPI; 2001-025136/03.
 XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
 PT inhibit angiogenesis in the treatment of disorders such as cancer,
 PT rheumatoid arthritis and psoriasis -
 XX Claim 7; Pages 543-546; 768pp; English.
 XX The present invention relates to human METH1 and METH2, (ME for
 CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
 CC The present sequence is an expressed sequence tag (EST) for METH. METH
 CC can be used for inhibiting angiogenesis in an individual, and for
 CC treating cancer, benign tumours, an ocular angiogenic disease,
 CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
 CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
 CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
 CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
 CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or
 CC atherosclerosis. METH can also be used in birth control. METH can also
 CC be used in diagnostic methods for the prognosis of cancer.
 XX Sequence 4180 BP; 1078 A; 1051 C; 1146 G; 905 T; 0 other;
 SQ
 Query Match 22.3%; Score 636; DB 22; Length 4180;
 Best Local Similarity 60.7%; Pred. No. 1.3e-122;
 Matches 1098; Conservative 0; Mismatches 700; Indels 12; Gaps 3;
 QY 631 AAGGTTTCGTGCTATCCCGCGGTAGTGGAGACGCTGGTGTGCGGACGAGTCAATG 690
 DB 1143 AAGCGATTTGTGTCACGCCCGGTTATGTGTGGAACCAATGCTGTAGTGCACCTCAATG 1202
 QY 691 GTCAAGTTCACGGCGCGGAGTGGAAATATCTGCTGACGCTGTGGCAACGCGGCG 750
 DB 1203 GCCGACTTCCAGCGCAGCGGTCTAAAGCATTAAGCTTCAACCCGTCTCGGTGGCAGCC 1262
 QY 751 CGACTTACCGCCATCCAGCATCTCAACCCCATCAACATGCTGTGTGTCAGGTGCTG 810
 DB 1263 AGGTTTACAGCATCCCGACATAGGAATTCATTAAGCTGTGTGTGTGGAAGATCTTG 1322
 QY 811 CTTCCTTAGAGATCGTGAATCCCGGCCCAAGTCAACGGCAATCGGCGCTGACGCTGCGC 870
 DB 1323 GTCATATACAGGAGCAGAGGACAGAGAGTACCTTCAATGCAGCTCTCACCCCTGG 1382
 QY 871 AACTTCTGTGCTGGCAGAGAGAGTGAACAAAGTGAAGTGAACAGCACCACCGGAGTACTGG 930
 DB 1383 AATTTCTGACGTGGCAGAGAACACACACAGCCCAAGTCAACCGGATCCAGAGCACTAT 1442
 QY 931 GACACTGCCATCTTTCACCGAGCAGGACCTGTGTGGAGCCACCACTGTGACACCCCTG 990
 DB 1443 GACATCGCAATCTGTTCACAGACAGAGATTTATGTGGCTCCCAACGCTGTGACACTCTC 1502
 QY 991 GGCATGCTGATGTGGGTGACCATGTGTGACCCCAAGAGAGCTGCTGTGTCATTGAGGAC 1050
 DB 1503 GGAATGGCAGATGTGGAACCGTATGTGACCCAGAGAGAGCTGCTGAGTCAATAGAGAT 1562
 QY 1051 GATGGGCTTCCATCAGCTTTCACACTGCCCACAGAGCTGGGCCACGCTGTTTCAACATGCC 1110
 DB 1563 GATGGTTTGAAGCGCCTTCCACACAGCCCATGAATTTGGGCCATGTGTTTAAACATGCCG 1622
 QY 1111 CATGACATGTGAAGTCTGTGAGGAGGTGTTGGGAGCTCCGAGAGCTCCGACCAACCATGATG 1170
 DB 1623 CACATGATGTGAAGCACTGTGCGAGCTTGAATGGTGTGAGTGGCGAATTCATCATGTATG 1682
 QY 1171 TCCCGGACCTCATCCAGATCGACCGGTGCCAACCCTTGGTGCAGCTGCAGTGCATGCATC 1230

DB 1683 GCCTCGATGCTCTCCAGCTTAGACCATAGCAGCCCTGGTGCACCTTGCAGTGCCTACATG 1742
 QY 1231 ATCACCGACTTCTGGACAGCGGCGACCGGTGACTCCCTCTGGACCAACCCAGCAGCC 1290
 DB 1743 GTCAGTCTTCTTAGATTAATGACAGCGGGAATGTTTGTATGGACAAAGCCCAAGATCCA 1802
 QY 1291 ATCTCCCTGCGGAGGATCTGCGGCGGCCAGCTTACACCCTGAGCCAGCAGTGGAGCTG 1350
 DB 1803 ATCAAGCTCCCTTCTGATCTTCCCGGTACCTTGTACGATGCCAACCCGCAATGTGCTGTT 1862
 QY 1351 GCTTTTGGTGGGCTCCCAAGCCCTGCTCTTACATGCAATGAC---TGCACCAAGCTCTGG 1407
 DB 1863 ACATTGCGAGGAGAAATCCAGCAGCTGCCCTGATGAGCCAGCAGATGTACTACTCTGTGG 1922
 QY 1408 TGCACCGGAGGAGGAGGAGGAGAGTGTGTCAGACCCGCCACTTCCCTGGGCGCAT 1467
 DB 1923 TGCAGTGGCAGCTCGGCTGGCTTACTGTGTGTCACCAACAAACACTTCCCTTGGCAGAT 1982
 QY 1468 GGCACAGCTGTGGCGGAGGAGGAGTCTGCTCTCAAGGGGCTGCGTGGAGAGACACAC 1527
 DB 1983 GGCACAGCTGTGGAGAGGAGGAGTGTGTGTGTCAGTGGCAAGTGCCTGGAACAGACAGAC 2042
 QY 1528 CTCACAAAGCAGCAG-----GGTGTGATGCTTCTGGGCGCAATGGATCCCTATGGCCCC 1581
 DB 2043 ATGAGCATTTTGTCTACTCTCTTTCATGGAAGCTGGGAGCCATGGGACCTGGGAGAC 2102
 QY 1582 TGCTCGCGACATGTGGTGGGCGGTGAGCTGCGCAGAGGAGTGCACCAACCCAC 1641
 DB 2103 TGCTCAAGAAACCTGTGGTGGTGGATTTCAATACAAATGAGAGATGTGACAAACCCAGTC 2162
 QY 1642 CCTGCCAACGGGCGAGCTACTGCGAGGAGTGGAGGTGAAATACCGATCTGCAACCTG 1701
 DB 2163 CCAAGAGACGGAGGAGTACTGTGAAGGAAACAGTCCCTACAGTCTCTGTAAACATC 2222
 QY 1702 GAGCCTCTCCCGCAGCTCAGCCCTCCGGAAGAGCTTCCGGGAGGAGAGTGTGAGCGTTTC 1761
 DB 2223 GAGGACTGTCCAGACAAATAA---CGGAAAAAGTTTCAGAGAGGAGCAGTGCAGGCGCAC 2279
 QY 1762 AACGGGTACACCAACAGCAGCACCACCGGCTACTCTGCGCGTGGCATGGTGCCTCAAGTAC 1821
 DB 2280 AATGAGTTTCCAAAGCTTCTCTTGGGAATGAGCCCACTAGATGAGCACCACCAAGTAC 2339
 QY 1822 TCCGCGTGTCTCCCGGAGCAAGTCAAGCTCATCTCCGAGGCAATGGCACTGGCTAC 1881
 DB 2340 GCGGCGTCTCCCAAGAGAGAGTGCAGCTCACTGTGAAGCCAAAGGCAATGGCTAC 2399
 QY 1882 TTCTATGTGCTGGCAGCAGGAGTGTGAGCAGGAGTGTGCTCTCTGCTGCTGCTGCTGCT 1941
 DB 2400 TTTTTCGCTTACAGCCCAAGGTTGTAGATGGCAGCTCCCTGTGTAGTCCAGACTCTACCTCT 2459
 QY 1942 GTCTGTGTCAGGCAAGTGCATCAAGCTGGCTGTGATGGGAACCTGGGCTCCGAAGAAG 2001
 DB 2460 GTCTGTGTCAGGCGAGTGTGTGAAGCTGGCTGTGATCGCATCATAGATTCACAAAG 2519
 QY 2002 AGATTGCAAGTGTGGGTGTGTGGGAGAGACAAATAGAGCTGCAAGAAAGTGAAGTGA 2061
 DB 2520 AAGTTTGAATGTGGCTTGTGGAGGAACAGGTTCCACATGCAAGAGATGTCAGGA 2579
 QY 2062 CTCCTCACCAAGCCCATGATGGCTTACAAATTTGTTGGTGGGCAATCCCGCAGGCGCTCA 2121
 DB 2580 ATAGTCACTAGTACAGACCTGGGTATCATGACATTTGTACAAATTCCTGCTGGAGCCACC 2639
 QY 2122 AGCATGCAATCCCGCAGGCGGTTTACAAAGGCTGTATCGGGGATGCAACACTACCTGGCT 2181
 DB 2640 AACATGGAAGTGAACATCGGAATCAAGGGGGTTCAGAAACAAATGGCAGCTTCTGGCT 2699
 QY 2182 CTGAAGAACAGCAAGGAGTACCTGCTCAACGGGCTTTCGTTGCTGCGCGGTGGAG 2241
 DB 2700 ATTAGAGCGCTGTGATGCTACCTATATTCTGATGGAACACTTCACTCTGCTCCACACTAG 2759
 QY 2242 CGGGAGCTGGTGTGAAGGCGAGTGTGCTGCGGTACAGCGGCGAGGCGACAGCGGTGGAG 2301
 DB 2760 CAAGACCTCACCTACAAAGGTAAGTGTCTTAAGGTACAGTGGTTCCTCGGCTGCGTGGAA 2819

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2003, 19:29:16 ; Search time 103 seconds
(without alignments)
8494.649 Million cell updates/sec

Title: US-10-009-332-2
Perfect score: 2853
Sequence: 1 atgctttgctgggcatcct.....gcgtctgagcgctgctga 2853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2.6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2.6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2.6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2.6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2.6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	628.2	22.0	4676	4 US-09-130-491-1	Sequence 1, Appli
2	618.8	21.7	4858	4 US-09-392-184-1	Sequence 1, Appli
3	613.3	21.5	3708	4 US-09-484-970B-58	Sequence 58, Appli
4	599.8	21.0	3126	4 US-09-392-184-7	Sequence 7, Appli
5	592	20.8	3638	4 US-09-369-364A-8	Sequence 8, Appli
6	584.8	20.5	4192	4 US-09-122-126B-1	Sequence 1, Appli
7	547.4	19.2	3250	4 US-09-122-126B-14	Sequence 14, Appli
8	545.8	19.1	3002	4 US-09-369-364A-1	Sequence 1, Appli
9	471.4	16.5	2825	4 US-09-369-364A-14	Sequence 14, Appli
10	411.2	14.4	5804	4 US-09-369-364A-12	Sequence 12, Appli
11	318	11.1	2114	4 US-09-130-491-7	Sequence 7, Appli
12	259.4	9.1	739	4 US-09-369-364A-10	Sequence 10, Appli
13	247.2	8.7	5357	4 US-09-392-184-5	Sequence 5, Appli
14	240.2	8.4	3218	4 US-09-369-364A-6	Sequence 6, Appli
15	215.8	7.6	1520	4 US-09-369-364A-3	Sequence 3, Appli
16	214.4	7.5	3885	4 US-09-369-364A-16	Sequence 16, Appli
17	187	6.6	2450	4 US-09-491-522-2	Sequence 2, Appli
18	187	6.6	6692	4 US-09-491-522-9	Sequence 9, Appli
19	179.4	6.3	2450	4 US-09-491-522-9	Sequence 9, Appli
20	179.4	6.3	4580	4 US-09-491-522-8	Sequence 8, Appli
21	162.2	5.7	703	4 US-09-392-184-6	Sequence 6, Appli
22	117.6	4.1	3675	4 US-09-930-872-3	Sequence 3, Appli
23	117.6	4.1	4042	4 US-09-930-872-5	Sequence 5, Appli
24	91.6	3.2	2848	4 US-09-369-364A-4	Sequence 4, Appli
25	79	2.8	2023	4 US-09-491-522-6	Sequence 6, Appli
26	77	2.7	385	4 US-09-392-184-23	Sequence 23, Appli
27	67.6	2.4	349	4 US-09-392-184-24	Sequence 24, Appli

28	61.4	2.2	1642	4 US-09-369-364A-18	Sequence 18, Appli
29	57	2.0	1476	4 US-09-930-872-1	Sequence 1, Appli
30	55.8	2.0	1803	4 US-09-369-364A-20	Sequence 20, Appli
31	50.4	1.8	657	3 US-08-985-526-2	Sequence 2, Appli
32	50.4	1.8	1326	3 US-08-985-526-4	Sequence 4, Appli
33	49.4	1.7	716	4 US-09-276-531-117	Sequence 117, App
34	47.4	1.7	3014	2 US-08-808-982-1	Sequence 1, Appli
35	47.4	1.7	3014	2 US-09-306-902A-1	Sequence 1, Appli
36	47.2	1.7	1173	3 US-08-706-216-5	Sequence 5, Appli
37	46.4	1.6	1101	4 US-09-152-060-23	Sequence 23, Appli
38	44.6	1.6	502	4 US-09-392-184-12	Sequence 12, Appli
39	44	1.5	7452	3 US-08-592-500-1	Sequence 1, Appli
40	44	1.5	7452	3 US-08-195-006-1	Sequence 1, Appli
41	44	1.5	7452	5 PCT-US94-07644A-1	Sequence 1, Appli
42	43.6	1.5	2063	1 US-08-715-554-1	Sequence 1, Appli
43	43.6	1.5	2063	2 US-08-583-118-1	Sequence 1, Appli
44	43.4	1.5	7218	1 US-08-232-463-14	Sequence 14, Appli
45	43	1.5	20235	1 US-07-642-734C-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-130-491-1
; Sequence 1, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodman, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130.491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054, 961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4676
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (460)....(3360)
US-09-130-491-1

Query Match	22.0%	Score 628.2;	DB 4;	Length 4676;
Best Local Similarity	57.1%	Pred. No. 1.1e-127;		
Matches 1305;	Conservative	0;	Mismatches 913;	Indels 69; Gaps 6;
QY	175	GCATTTCAGGAGGACTTTTACCTACACCTGAGCGCGGATGCTCAGTCTTGCGTCCGCC	234	
DB	697	GCCTTTACCAAGAGCTGGATCTGGAGCTGCGGCCGACAGCAGCTTTTTCGCCCGCCGC	756	
QY	235	TTCTCACTGAGCATCTGGGGTCCCTCCAGGGGCTCCAGGGGGCTC---TTCAGAC	291	
DB	757	TTCAGCGCTCCAGAGCTGGGGCGCAATCCGGGTCGAGAGACCGCTTCCGGAAACCGAC	816	
QY	292	CTGGAGCGTCTTCTATTCTTGGGAGCGTGAACGCCGCGGAGCTCGTTCGCTGCTGTG	351	
DB	817	CTGGCGCACTGCTTCTACTCCGGCACCGTGAATGCGATCCAGCTCGCTCGCGCCCTC	876	
QY	352	AGCCTGTGGGGGCGCTCCGGGAGCCTTTGGCTACCGAGGCGCGGAGTATGCTATTAGC	411	
DB	877	AGCCTCTCGAGGGCGCTCGCGCGCCCTTCTACCTGCTGGGGGAGGCGTATTTTATCCAG	936	
QY	412	CGCTGCGCAATGCTAGCGCGC-CGGCGGCGCAGCGCAACAGCGGCGGACACCTTCT	470	
DB	937	CGCTGCGCGCGCGCAGCGCGCTCCACCGCGCCCGCCCGGAGAACCCCGGCA	996	

QY 471 CAGCCGCGGGGTGTTCCGGGGGGGGCTTCCGGAGACCCCACTCTCGTGGGGGGTGGC 530
 Db 997 CCACTACAGTTCCACCTCTCGGGGGGAATGCGAGCGGCGACGTAGCGGCGACGTGGCG 1056
 QY 531 CTGGGCTGGAACCCCGCCATCTAGGGCCCTGGACCTTACAGCCGCGCGGGGG 590
 Db 1057 GTCTGGAGCAGAGCCCGCGGCTGGAAGAGGAGACCGAAGAGAGGAGGAAAGG 1116
 QY 591 CTTTCGGGAGAGTCGTAGCGGCGAGGCTCAGTGGTCCGCGAGGACCCGCGACTGCAAGCGTA 627
 Db 1117 ACTGAGGCGAGGACGAAGGCTCAGTGGTCCGCGAGGACCCGCGACTGCAAGCGTA 1176
 QY 628 -----GCCAAGCGTTTCGTGTCTATCCCGGGTAC 657
 Db 1177 GGAGAGCCACAGAACTGGAAGCATAAGAAAGAGGATTTGTCTCAGTCAACGCTAT 1236
 QY 658 GTGGAGACGCTGGTGTCCGGGAGGAGTCAATGGTCAAGTTCAGCGGCGCGACCTGNA 717
 Db 1237 GTGGAACCATGCTGTGGCAGACCATGATGGGAGAAATCCAGCGGAGTGGTCTAAG 1296
 QY 718 CATTATCTCTCAGCTGTGGCAAGCGGCGGCGACTCTACCGGCATCCCGAGCATCTC 777
 Db 1297 CATTACCTTCTCAGCTGTGGTGGCAGCAGATTGTCAAAACCCCGAGCATTCGT 1356
 QY 778 ACCCCATCAACATCTTGTGTCAAGTGTCTTCTTTAGAGATCGTACTCCGGGCC 837
 Db 1357 AATTCAGTTAGCTGTGGTGGTGAAGATCTTGGTCACTCCAGATGAACAGAGGGCGC 1416
 QY 838 AAGTCCAGCGGAATCGCGCCCTGAGCTGCGCAACTCTGTGCTGGCAGAGAGCTG 897
 Db 1417 GAAGTCACTTCAAGTCTGCCCTCACTTGGCGAATTTTGAACCTGGCAGAGCAGCAC 1476
 QY 898 AACAAAGTGTGACAGCACCCGAGTACTGGGACTGCGCATCTCTTACACGAGCAG 957
 Db 1477 AACCACCCAGTACCGGGATGACAGGACTATGACAGCAATTTTTCACAGACAG 1536
 QY 958 GAGCTGTGGAGCCACACACTGTGACACCTGGGCGATGGCTGATGGGTACCATGTGT 1017
 Db 1537 GACTGTGTGGTGGTCCAGACATGTGATCTTGGGATGGCTGTGTTGGAATGTGTGT 1596
 QY 1018 GACCCCAAGAGAGTGTCTGTATTGAGAGCATGGGCTTCCATCAGCTTCCACACT 1077
 Db 1597 GATCCAGCAGAGTGTCTCCGTCATAGAGATGATGTTTACAAGCTGCGCTTCCACA 1656
 QY 1078 GACCAGAGCTGGGCGACGTGTTCAACATGCCCCATGACAAATGTGAAAGTCTGTGAGGAG 1137
 Db 1657 GCCATGAATTAGGCGACGCTTTAATATGCCACATGATGATGACAAAGCAGTGTGCCAGC 1716
 QY 1138 GTGTTTGGAGCTCGGACCAACACATGATGTCGCCGACCTCATCCAGATCCAGCT 1197
 Db 1717 CTTAATGGTGTGAACCCAGGATTTCCACATGATGGGCTCAATGCTTCCAACTGGACAC 1776
 QY 1198 GCAACCCCTGTGAGCTGCAAGTGTGCTGATCATCATCAGCTTCTCGACAGCGGGCAQ 1257
 Db 1777 AGCCAGCTTGTCTCTTGGAGTGCCTACATGATTAATCTTCTGGATAATGGTCTAT 1836
 QY 1258 GTGACTGTCTCTGAGCAACCCAGAGAGCCATCTCCCTGCCGAGGATGTGCCGGGC 1317
 Db 1837 GGGGAATGTTGATGGACAAGCCTCAGAATCCATACAGTCCAGCGGATCTCCCTGGC 1896
 QY 1318 GCCAGTACACCTGAGCCAGCAGTGGAGCTGGCTTTTGGCGTGGGCTCCAGCCCTGT 1377
 Db 1897 ACTCGTAGGATGCCAACCGGAGTGGCAGTTTACATTTGGGGAGGAGCTCCAAACACTGC 1956
 QY 1378 CTTTACATCAGTAC-----TGCAACCAAGTGTGGTGCACCGGGAAGGCCAAGGACAGATG 1434
 Db 1957 CCGGATGACGACGACATGATAGACCTTGTGGTGTACCGGACCTCTGTGGGTGTGTG 2016
 QY 1435 GTGTGCCAGACCCGACATTCCTCCCTGGGCGGATGACACAGCTGTGGGAGGCGCAAGTCT 1494
 Db 2017 GTGTGTCAAAACCAACACTTCCCGTGGCGGATGGCAGCTGTGGAGAGGAAATGG 2076
 QY 1495 TGCCCTAAAGGGGCTGCGTGGAGAGACACACACTCAACAG-----CACAGGTGGAT 1548

Db 2077 TGTATCAACGCAAGTGTGTGAACAAAACCCACAGAAGCAATTTTGATACGCCCTTTTCAT 2136
 QY 1549 GGTTCCTGGGCAAAATGGGATCCCTATGGCCCTGTCTCGCGCACATGTGTGGGGGGTG 1608
 Db 2137 GGAAGCTGGGGAATGTGGGGCCCTTGGGAGAGATGTTTCGAGAACGTCGGGTGGAGGATC 2196
 QY 1609 CAGTGGCCAGGAGGAGTGCACCAACCCACCCCTGCCAACCGGGGCAAGTACTGCGAG 1668
 Db 2197 CAGTACAGTATGAGGGAATGTGACAAACCCAGTCCCAAGAAATGGAGGAAATGACTGTGAA 2256
 QY 1669 GAGTGTAGGTTGAATACCGATCTGCAACCTGGAGCCCTGCCCGACCTCAGCCTCGGA 1728
 Db 2257 GGCAAAGAGTGGCTACAGATCCCTGTAACCTTGGAGACTGTCCAGAC---AATAATGA 2313
 QY 1729 AAGAGCTTCGGGAGGAGCAGTGTGAGCTTCAACCGCTTACAACACACACACCAACCGG 1788
 Db 2314 AAAACCTTTAGAGAGGAACATGTGAAGCACACAAACGATTTTCMAAGCTTCTTTGGG 2373
 QY 1789 CTCATCTCGCCGTGGATGGGTGCCCAAGTACTTCCCGCGTGTCTCCCGGAGCAAGTGC 1848
 Db 2374 AGTGGCTCGGGTGAATGATTTCCCAAGTACGCTGGCGTCTCACCAAAAGGACAGTGC 2433
 QY 1849 AAGCTCATCTGCCGAGCAATGGCACTGGCTACTTCTATGTGTGGCACCCCAAGTGGTG 1908
 Db 2434 AAGCTCATCTGCCAAGCAAGGATTTGGTACTTCTTCTGCTTTCGACGCCCAAGGTGTA 2493
 QY 1909 GACGGCAGCTGTCTCTCTGCTGACTCCACCTTCCGCTGTGTCCAAAGCAAGTGCATCAAG 1968
 Db 2494 GATGGTACTCCATGTACCCAGATTTCCACCTCTGTCTGTGCAAGGACASTGTGTA 2553
 QY 1969 GCTGGCTGTGATGGGAACCTTGGGCTCCAAAGAGAGATTCGACAAAGTGTGGGGTGTGGG 2028
 Db 2554 GCTGTTGTGATGATCATAGACTTCCAAAGAAAGTTTGATAAATGTGTGTGCGGG 2613
 QY 2029 GGACACATFAGAGCTGCAAGAGGTCAGTGGACTCTTCCACCAAGCCCATGCAATGCTAC 2088
 Db 2614 GGAATGGATCTACTTGTAAAAAATATCAGGATCAGTTACTAGTCAAAACCTGGATAT 2673
 QY 2089 AATTTCTGTGTGGCTATCCCGGAGCGCTCAAGCATCGACATCCGCCAGCGGTTAC 2148
 Db 2674 CATGATATCATCAATTCACACTGGAGCCACACACATCGAAGTGAACAGCGGAACCCAG 2733
 QY 2149 AAGGGTGTATCGGGGATGACACTACTGGCTGTGAAGAACAGCCCAAGCAAGTACCTG 2208
 Db 2734 AGGGATCCAGGAACAATGGCAGCTTTTGGCCATCAAGCTGTGATGSCATATATT 2793
 QY 2209 CTCACCGGCAATTCGTGTGTGGGCTGGAGCGGAGCTGGTGGTGAAGGCGAGTCTG 2268
 Db 2794 CTTAATGGTACTACATTTGTCCACCTTAGACAAGACATTTATGACAAGGTGTGTGTC 2853
 QY 2269 CTGGGTACAGCGGCGCACGCGGTGGAGCGCTGAGAGCTTCCCGGCCCATCCTG 2328
 Db 2854 TTGAGGTACAGCGCTCTCTCGGCAATTTGAAAGATTCGAGCTTTAGCCCTCTCAA 2913
 QY 2329 GAGCCGTGACCGTGGAGGTCCTCTCGTGGGAAGATGACACCGCCCGGGTCCCGTAC 2388
 Db 2914 GAGCCCTTGACCATCCAGTCTTACTGTGGCAATGCCCTTCGACCTAAATTAATAC 2973
 QY 2389 TCCTTCT 2395
 Db 2974 ACCTACT 2980

RESULT 2
 US-09-392-184-1
 : Sequence 1, Application US/09392184
 : Patent No. 6395889
 : GENERAL INFORMATION:
 : APPLICANT: Robison, Keith E.
 : TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN
 : TITLE OF INVENTION: PROTEASE HOMOLOGS
 : FILE REFERENCE: 5800-55

; CURRENT APPLICATION NUMBER: US/09/392,184
; CURRENT FILING DATE: 1999-09-09
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1

; LENGTH: 4858

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc.feature

; LOCATION: (1)...(4858)

; OTHER INFORMATION: reprotolysin (ADAM family of metalloprotease)

; NAME/KEY: misc.feature

; LOCATION: (1)...(4858)

; OTHER INFORMATION: n = A,T,C or G

US-09-392,184-1

Query Match 21.7%; Score 618.8; DB 4; Length 4858;
Best Local Similarity 57.1%; Pred. No. 1.3e-125;
Matches 1306; Conservative 0; Mismatches 912; Indels 70; Gaps 7;

```
QY 175 GCATTTCAGGAGGACTTTTACCTACACCTGACGCGCGATGCTCAGTTCCTGGCTCCCGCC 234
DB 691 GCCTTTGACGACGAGTGGATCTGGAGCTGCGCGCCGACAGAGCTTTTGGCGCCCGCC 750

QY 235 TTCTCCACTGAGCATCTGGCGCTCCCGCTCCAGGGCTCACCAGGGGCTC---TTCAGAC 291
DB 751 TTCACGCTCCAGAACCTGGGGCGCAAAATCCGGTCCGAGACGCGCTTCCGGAAACCGAC 810

QY 292 CTGGAGCGCTCTTATCTGGGAGCTGTAAGCCGCGAGCGGACGCTGCTGCTGCTG 351
DB 811 CTGGCGAGCTCTTACTCTCGGACCGGTAATGGCGATCCGAGCTCGGCTCGCGCCCTC 870

QY 352 AGCCTGTGCGGGGGCTCCGCGAGCGCTTGGCTACCGAGGCGCGAGTATGCTATTAGC 411
DB 871 AGCCTGTGCGAGGCGTGGCGCGCTTCTACTGCTGGGAGCGGATTTTCATCCAG 930

QY 412 CCCTGCGCCATGCTAGCGGC--CGCGCGCGAGCGCAACAGCGGCGCACCTTTCT 470
DB 931 CCCTGCGCGCGCGCGAGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCA 990

QY 471 CCAGCGCGGGGTGTTCCGGGGGGCTTCCGAGACCCACCTCTCGCTGCGGGGTGC 530
DB 991 CCCTACAGTTCACCTCTCTGGCGGGAATCGGAGGCGAGCTAGCGGCGACGTGCGGG 1050

QY 531 CTCGGCTGGAACCCCGCCATCTACGGCCCTTGGACCTTACAGCCCGCGCGCGCGG 590
DB 1051 GTCGTGAGCAGCAGCGCGCGCTGGGAAAGCGGAGACCGAGAGCGAGCAAGG 1110

QY 591 CTTGGGGAGAGCTGTAGCGGGGCGAGTCTGGGCGC----- 627
DB 1111 ACTGAGGCGGAGGAGCAAGGGGCTCAGTGGTCCGCGCAGGACCCGCGCACTGCAAGGCGTA 1170

QY 628 -----GCCAAGCGTTTCGTCTATCCCGGGTAC 657
DB 1171 GGACGCCACAGGAAGTGAAGATCAAGAGAGCGATTTGTGTCAGTCACCGCTAT 1230

QY 658 GTGGAGCGCTGGTGTGCGGAGGATCAATGGTCAAGTTCAGGGCGCGGACCTGGAA 717
DB 1231 GTGGAACCATGCTTGTGCGAGACCAAGTCGATGGCAGAAATCCACGCGAGTGGTCTAAG 1290

QY 718 CATTAATCTCT--GACGCTCTGGCAACGCGCGGCGGAGCTCTACCGCCATCCAGCATCT 776
DB 1291 CATTAATCTCTCAACGTTGTTTTCGGTGGCAGCCAGATTTGTACAAACACCCCGCATTCG 1350

QY 777 CAACCCCATCAACATGTTGGTCAAGGTGCTGCTTCTTAGAGATCTGATCTCGGCGC 836
DB 1351 TAATTACGTTAGCTGGTGGTGAAGATTTGGTCAATCCAGTGAACAGAGAGGGGCC 1410

QY 837 CAAGGTCAACGCAATGGGCGCTGACGCTGGCAACTTCTGTGCTGGCAGAGAGCT 896
DB 1411 GGAAGTGACCTCAATGCTGCCCTCCTCCTGCTGGGAACTTTTGAACACTGGCAGACCA 1470
```

```
QY 897 GAACAAAGTGAAGTGAACAGCACCCCGAGTACTGGGACACTGCCATCTCTTTCACAGCA 956
DB 1471 CAACCCACCGAGTGAACCGGATGACAGCACTATGACAGCAATTTCTTTTCACAGCA 1530

QY 957 GGACCTGTGTGGAGCCACACCTGTGACACCTGGGCTGATGTGGG:ACCATGTG 1016
DB 1531 GGACTTGTGTGGTCCAGACATGTGATCTCTGGGATGGCTGATGTGGAAGTGTGTG 1590

QY 1017 TGACCCCAAGAGAAGTCTCTCTGTCATTGAGAGCATGGGCTTCCATCAGCTTTCACAC 1076
DB 1591 TGATCCGAGCAGAGCTGCTCGCTCATAGAAAGATGGTTTACAAGCTGCTTTCACAC 1650

QY 1077 TGCCCAACAGCTGGGCGAGCTGTTCAACATGCCCATCACAATGTGAAGTCTGTGAGGA 1136
DB 1651 AGCCCATGAATTAGCCACAGTGTTTAATCATGCCATCATGATGCAAGAGTGTGCCAG 1710

QY 1137 GGTGTTTGGGAAGTCCGAGCCAAACACATGATGTCCCGGACCTCATFCCAGATCGACCG 1196
DB 1711 CTTAATGTTGTAACAGGATTTCCACATGATGGCGTCAATGCTTTCCAACTGGACCA 1770

QY 1197 TGCCCAACCCCTGGTCAAGCTGCTGCTGATGCTGCTGATCATCACCAGCTTCTTGAGCAGCGGCA 1256
DB 1771 CAGCCAGCGCTTGGTCTCTCTTGCAGTGCCTACATGATTTACATCATTTCTGGATAATGTGCA 1830

QY 1257 CGGTGACTGCTCTCTGGAGCAACCCAGCAAGCCCATCTCCCTGCCGAGGATCTGCCGGG 1316
DB 1831 TGGGAATGTTGATGGAGCAAGCTCAGAAATCCCATACAGCTCCAGCGGATCTCCCTGG 1890

QY 1317 CGCCAGCTACACCTTGAGCCAGCAGTGGAGCTGGCTTTTGGCGTGGGCTCCAAAGCCCTG 1376
DB 1891 CACCTCGTACGATGCCAACCGGCGAGTGCAGTTTACATTTGGGGAGGACTCCAAACACTG 1950

QY 1377 TCCTTACATGCAATAC---TGCACCAAGCTGTGTGTCACCGGAGGCAAGGACAGAT 1433
DB 1951 CCCTGATGACGACAGCAATATGATGACCTTGTGGTGTACCGGACCTCTGTGTGGGGTCT 2010

QY 1434 GGTGTGCCAGACCCCGCCATCTCCCTTGGCGGATGGACACAGCTGTGGCGAGGGCAAGCT 1493
DB 2011 GGTGTGTCAAACCAACACACTTCCCGTGGCGGATGGCAGCAGCTGTGGAGAGGGAATG 2070

QY 1494 CTGCTCAAGGGGCGCTGCTGGAGAGACACAACTCAACAG-----CACAGGTGGA 1547
DB 2071 GTGTATCAACGGCAAGTGTGTGAACAAACGACAGAAAGCATTTTGATACGCTTTTCA 2130

QY 1548 TGGTTCCTGCGCCAAATCCGATCCCTATGCGCCCTGCTCGCGCACATGTGCTGGGGCGT 1607
DB 2131 TGGAGCTGGGATGTGGGGCGCTTGGGAGAGCTGTTCGAGAACGTCGCGTGGAGAGT 2190

QY 1608 GAGCTGGCCAGGAGGAGTGCACCAACCCCGCTGCCAACGGGGGCAAGTCTGCGGA 1667
DB 2191 CCAGTACACGATGAGGGAATGTGACAAACCCAGTCCCAAGAAATGAGGGAAGTACTGTA 2250

QY 1668 GGGAGTGAAGGTGAATACCATCCTGCAACCTGAGCGCTTGCACAGCTTCAGCTCCGG 1727
DB 2251 AGGCAAGAGTGGCGCTACAGATCCTGTAACCTTGGAGGACTGTCCAGAC---AA-TAATGG 2307

QY 1728 AAAGAGCTTCCGGGAGGAGCAGTGTGAGGCTTTCACAGGCTTACAAACACAGCACAACCG 1787
DB 2308 AAAACCTTTAGAGAGGACAAATGTGAGCAGACACAGAGTTTTCAAAAGCTTCTTTGG 2367

QY 1788 GCTCACTCTCGCGGTGCGATGGGTGCCCAAGTACTCCGGGCTGTCTCCCGGGCAAGT 1847
DB 2368 GAGTGGGCGCTCGGTGGAATGGAATCCCAAGTACGCTGCGCTCTCAACCAAGGACAGTG 2427

QY 1848 CAAGCTCATCTGCGGAGCCATGCACTGGCTACTTCTATGTGCTGCGACCCCAAGGTGGT 1907
DB 2428 CAAGCTCATCTGCGCAAGCAAGGCAATGGCTACTTCTCTCTTTTTCAGGCCCAAGGTGT 2487

QY 1908 GGACGGCAGCGTGTCTCTCTGCTACTCCACTCGCTGTGTGTGTCAGGCAAGTGCATCAA 1967
DB 2488 AGATGCTACTCCATGTAGGCCAGATTCACCTCTGTCTGTGTGCAAGGACAGTGTGTA 2547

QY 1968 GGTGGCTGTGATGGGAACCTGGGCTCCAAAGAGAGATTCACAAAGTGTGGGTGTGTG 2027
```


QY 1882 TTCTATGTGCTGACACCAAGGTGGTGAGCGGACGCTGTGCTCTCTGACTCCACCTCC 1941
Db 1440 TTCTTGTGTTTGCAGCCCAAGGTGTAGATGTACTTCAATGATAGCCAGATTCACCTCT 1499
QY 1942 GTCTGTGTCACAGGCAAGTGCATCAAGGCTGGTGTGATGGAACTGGGCTCCCAAGAG 2001
Db 1500 GTCTGTGTCACAGGCAAGTGTGTAAGGCTGGTGTGATCGCATCATAGCTCCCAAAAAG 1559
QY 2002 AGATTCCGACAAGTGTGGGTGTGTGGGGGAGACAATAAGAGCTGCAAGAGGTGACTGGA 2061
Db 1560 AGATTTCGATAATGTGGTGTGGGGGGAATGGATCTACTTGTAAAAAATAATACAGGA 1619
QY 2062 CTCTTCCACCAAGCCCATGATGCTACAAATTCGTGGTGGCCATCCCGCAGGCGCTCA 2121
Db 1620 TCAGTTACTAGTCAAAACCTGATATCATATCATCAATTCACACTGGAGCCACC 1679
QY 2122 AGCATCGACATCCGCCAGCGGTGTACAAGGCTGATCGGGAT-GACAACTACTGCG 2180
Db 1680 ACATCGAAGTGAACAGCGGAACACAGAGGGATCCAGGAACAATGGGCGAGCTTCTTTCG 1739
QY 2181 TCTGAAGAACAGCGCAAGCAAGTACTGCTCAACGGGCAATTCGTGGTGTGGCGGTGGA 2240
Db 1740 CATCAAGCTGTGATGGCAATATATCTTAATGGTGACTACACTTTGTCCACCTTGA 1799
QY 2241 GCGGGACCTGGTGGTGAAGGGCAGTCTGCTGCGGTACAGCGCACGGCGGTGGA 2300
Db 1800 GCAAGACATATGATCAAAAGGTGTGCTTGAGGTACAGCGCTCTCTGCGGCATGGA 1859
QY 2301 GAGCCTGAGGCTTCCCGCCCATCTGAGCGGCTGACCGGTGAGGTCTCTCGTGGG 2360
Db 1860 AAGAATTCGACGCTTAGCCCTCTCAAGAGGCCCTGACCATCCAGGTTCTTACTGTGG 1919
QY 2361 GAAGATGACACGCGCCCGGCTCGGTACTCTCTTCT 2395
Db 1920 CAATGCCCTTGCACCTAAATTAATACACTACT 1954

RESULT 4
US-09-392-184-7/c
; Sequence 7, Application US/09392184
; Patent No. 6395889
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: 5800-55
; CURRENT APPLICATION NUMBER: US/09/392,184
; CURRENT FILING DATE: 1999-09-09
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 3126
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(3126)
; OTHER INFORMATION: repolysin (ADAM family of metalloprotease)

Query Match 21.0%; Score 599.8; DB 4; Length 3126;
Best Local Similarity 57.4%; Pred. No. 1.5e-121;
Matches 1304; Conservative 0; Mismatches 907; Indels 60; Gaps 10;
QY 157 CTCATTTTTCAGATCACAGCATTTTCAGGAGGACTTTTACCTACACTGACGCGGATGCT 216
Db 3122 CTCGGCTCCCACTGTGCGCTTCGGCAAGGGCTTCGTGCTGCGCTGGCGCCGCGAGC 3063
QY 217 CAGTTCTTGGCTCCGCTCTTCCACTGACGATCTGGCGCTTCCCTTCCAGGCGCTCAC 276
Db 3062 AGCTTCTTGGCGCGGACTTCAAGATCGAGCGCTTCGGGGCTCCGCGCGGGG---ACC 3006
QY 277 GGGGCTCTTCAGACCTCGCAGCTGCTTCTATTCTGGGACGTGAACGCGGAGCGGAC 336

Db 3005 GGGGCGAGCGGGGCTGCGCGGCTGTCTTCTCCGCGACCTCCACGGGTCGCGGAG 2946
QY 337 TCGTTCTGCTGTGAGCTGTGCGGGGCTCCGCGAGGCTTTTGCTACCGAGGCGCC 396
Db 2945 TCGTGGCGGGCTCAGCTGTGCGGGGCTGAGCGGCTCTTCTCTGTGAGCGGAG 2886
QY 397 GAGTATGTATAGCCGCTGCCCAATGCTAGCGCGCGGGCGGCGAGCGAACAGCCAG 456
Db 2885 GAGTTACATCCAGCGCGAGGCGCGGGGCTCCCTGGGTACGCGCACCGCTGCGAG 2826
QY 457 GCGCACACCTTCTCCAGCGCGGGGTGTTCCGGGGGGGCTTCCGAGAGACCCACCTCT 516
Db 2825 CGCTGGGGTCCCGCGGAGCGCCGCCCTCCCGGAGGACCCGAGTGGAGAGCG 2766
QY 517 CGCTGCGGGGTG-GCCTCGGGCTGGAACCGCCCATCTACGGGCGCTTGACCCCTTACAA 575
Db 2765 GGAGAGGTGTCAGAGCAGGAGAGAGAGACCAACAGGAGGACAGGAGGAGAGACCAA 2706
QY 576 GCCCGGGGGCGGGCTTCGGGGAGAGTGTGAGCGCGGCGAGGTCTGGGGCGG-----CC 630
Db 2705 GAAGAGGAGCAGAGGCGGTAGCGAGCGCGCACCGCCCTGGGGGCGCACGAGTAGGACC 2646
QY 631 AAGCGTTTCTGTCTATCCCGCGGTAGCTGGAGACGCTGTGTGCGGGGAGGAGTCAATG 690
Db 2645 AAGCGTTTGTGTCTGAGGCGCGCTTCTGTGAGACGCTGTGTGGCGGATGCGTCCATG 2586
QY 691 GTCAAGTTCCACGCGCGGACCTGGAACATATTCTGTCTGACGCTGTGTGGCAACGCGGCG 750
Db 2585 GTCCTCTTCTACGGGGCGGACTGCAGAACCACTCTGACGTTAATGTCTGTGGCGACC 2526
QY 751 CGACTCTACCGCCATCCAGCATCTCTCAACCCCATCAACATCGTTGTGTGTCAAAGTGTG 810
Db 2525 CGAATCTACAAGACACCCAGCATCAAGAAATTCATCAACCTGATGTGTGTAAGTGTG 2466
QY 811 CTCTTAGAGATCGTGACTCGGGGCCCAAGGTCAACGGGCAATGGGCCCTTGACGCTGCGC 870
Db 2465 ATCGTAGAGATGAAATAATGGGCGCCAGAGGTGTCCGACAAATGGGGGCTTACACTGCT 2406
QY 871 AACTTCTGTGCTGGCAGAGAAGCTGAACAAAGTGTGAGTGAACAGCAGCCCGAGTACTGG 930
Db 2405 AACTTCTGCACTGGCAGCGCGCTTTCACACGCCAGCGACCCACCCAGAGCACTAC 2346
QY 931 GACACTGCCATCTCTTTCACAGGAGGACCTGTGTGG-----AGCCACCACTGTGACACC 987
Db 2345 GACAGGCCATCTCTGTCTACACAGCAAACTTCTGTGGCAGGAGGCTGTGTGACACC 2286
QY 988 CTGGCATGCTGATGTGGGTACCATGTGACCCCAAGAGAGCTGCTCTGTCTATTGAG 1047
Db 2285 CTGGGTGCGACACATCGGGACCATTTGTGACCCCAACAAAGCTGCTCCGCTGATCGAG 2226
QY 1048 GACGATGGGCTTCCATCAGCCTTTCACACTGCCCAGAGCTGGGGCCACCTGTTTAAACATG 1107
Db 2225 GATGAGGGGCTCCAGCGGCCACACCTTGCGCCCATGAATAGGCGACCTCTCAGCATG 2166
QY 1108 CCCCATGACAAATGTGAAAGTCTGTGAGGAGGTGTTGGGAGCTCCGAGCGCAACACATG 1167
Db 2165 CCCCAGGACACTCCCAAGCCCTTCACACGCGCTCTTCGGGCCCATGGGCAAGCACCAGTG 2106
QY 1168 ATGTCCCGGACCTCATCCAGATCGACCGGTGCGCAACCCCTGTGTCAGCTGCTGCTGCC 1227
Db 2105 ATGGACCGGTGTGCTCCACCTGAACAGACGCTGCCCTGGTCCCTCTGCGAGCGCATG 2046
QY 1228 ATCATCCGACTTCTTGGACAGCGGCGCAGGTGACTGCTCTCTGGACCAACCCAGCAAG 1287
Db 2045 TATCTCACAGAGCTTCTTGAGGGGGGAC-----CGATTTCAGGTAT 2003
QY 1288 CCCATCTCCCTGCCAGGATCTGCGGGGCGGCGAGCTACACCTGAGCCAGCAGTGCAG 1347
Db 2002 TTATGGCCCTGTGCTAGTGTGTACAGCTTGCAAT-----ATGCTCTTTGGG 1955
QY 1348 CTGGCTTTTGGGCTGGGCTCCAGCCCTGCTTACATGCACTACTGACCAAGCTGTGG 1407

```

Db 1954 CCGGATTTCGGCCACATGCCCCAACACCTCTGCTCAGGAGCTCTGGCCCGCAGCTTTGGTGC 1895
QY 1408 TGCACCGGAAGCCCAAGGACAGATGGTGTGCAGACCCGCACTTCCCTGGGCGCAT 1467
Db 1894 CACACTGATGGGCTGAGCCCTGTGCCACAGAAATGGCAGCTGCGCTGGGCTGAC 1835
QY 1468 GGCACCAAGCTGTGGCGAGGCAAGCTCTGCTCAAAAGGCGCTGC-----GTGGAGAG 1520
Db 1834 GGCACCGCTGTGGGCGCTGGGCACTCTGCTCAGAAGGCAAGCTGTCTACCTGAGGAGGA 1775
QY 1521 ACACAACTCAACAGCAGAGGTGATGGTTCCTGGGCAAAATGGGATCCCTATGGCCC 1580
Db 1774 AGTGGAGAGCCCAAGCCCGCTGATAGAGGCTGGGACACCGTGGGACCTGGGGAGA 1715
QY 1581 CTGCTGCGCACATGTGTGGGGCGTGCAGCTGGCCAGGAGCAGTGCACCAACCCAC 1640
Db 1714 ATGTTCTCGGACCTGTGGAGGAGGTACAGTTTTCACACCTGTAGTGAAGACCCCGA 1655
QY 1641 CCTTGCACCGGGGCAAGTACTGCGAGGAGTGAAGGTGAATACCGATCTCTGCAACCT 1700
Db 1654 GCCTCAGAATGGAGGAATACCTGCTGGTGGAGAGCAAGTACCACTGATGCCACAC 1595
QY 1701 GGAGCCCTGCCAGCTACGCTCCGGAAGAGCTTCCGGAGGAGCAGTGTGAGCTTT 1760
Db 1594 GGAGGAATGCCCC-----CCTGACGGGAAAGCTTTCAGGAGCAGCAGTGTGAGAATGA 1541
QY 1761 CAACGGCTACACACAGCAGCACCAGCGCTCACTCTGCCGTGGCATGGTGCGCCCAAGTA 1820
Db 1540 TAATGCCTACAATTAACACTGACATGACATGGAGGGGAATCT---CCTCAGTGGTCCCAAGTA 1484
QY 1821 CTCGGGCGTGTCCCGGACAGATGCAAGCTCATCTGCGCGTGGCATGGTGCGCCACTGGCTA 1880
Db 1483 TGTGGGGTGTCCCGGAGCCGCTGCAAGTTGTTCTGCGGAGCCCGGGGAGGAGCGA 1424
QY 1881 CTTCTATGTGTGCGACCAAGGTGTGGAGCGGCTGTCTCTCTGACTCCACCTC 1940
Db 1423 GTTCAAGTGTTCGAGCCCAAGTGTATGGACCCCTGTGTGGGCCAGAAACACTGGC 1364
QY 1941 CGTGTGTCCAAAGCAAGTGTCAATGAAGCTGGCTGTGATGGGAACCTGGGCTCCAGAA 2000
Db 1363 CATCTGTGCTGGCCAGTGTGTCAAGGCGGCTGTGACCATGTGTGAGCTCGCTCG 1304
QY 2001 GAGATTCGACAAAGTGTGTGGGGAGACAATAAGAGCTGCAAGAGTGTGACTGG 2060
Db 1303 GAAGCTGGACAATAAGCGGGTGTGTGGGGCAAGGCAACTCTGCAGGAAGTGTCCGG 1244
QY 2061 ACTCTTCAACCAAGCCCATGATGCTACAATTTGCTGGTGGCCATCCCGCAGCGGCTC 2120
Db 1243 GTCCCTCACCCCCCACTAATATGGCTACATGACATGTCAACCATCCCACTGCTGGCAG 1184
QY 2121 AAGCATCGACATCCCGCAGCGGTTTACAAGGGCTGATCGGGGATGACAACTACCTGGC 2180
Db 1183 TAATATTGACGTGAAGCAGCGGAGCCACCGGGTGTGCAAGCACTGTTGAGCTCGCTCG 1124
QY 2181 TCTGAGACACCAAGCAGTACCTGCTCAAGGGCAATTTGCTGTGTGCGGGTGA 2240
Db 1123 GCTGAAGACGGCTGATGGCAGTACCTGCTCAAGCGCAACTGCGCCATCTTGCATAGA 1064
QY 2241 CGGGAGCTGTGTGAGGAGCAGTGTCTGCGGTACAGCGGCAAGCGGACAGCGGTGA 2300
Db 1063 GCAGGACATCTTGTGAAGGGAGCACTCTGAAGTACAGCGGCTCCATCCCACTCCGGA 1004
QY 2301 GAGCCTGAGGCTTCCCGGCGCCATCTTGGAGCCGCTGACCGTGGAGTCTCTCCGT--- 2357
Db 1003 CGGCTGAGAGCTTCCGGGCTTGGCAGAGCCCTTGACAGCTGAGCTCTGACAGTCCC 944
QY 2358 GGGGAGATGACACGGCCCGGCTCCGTACTCTCTTCTATCTGCCCAAGA 2408
Db 943 TGGGAGGTCTTCCCGGCAAAAGTCAAAATACACCTTTTGTCTTCTAATGA 893

```

RESULT 5
US-09-369-364A-8

```

; Sequence 8, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 3638
; TYPE: DNA
; ORGANISM: Mus musculus ADAMTS-8
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (278)..(2992)
; NAME/KEY: misc_feature
; LOCATION: (3636)
; OTHER INFORMATION: n = T
US-09-369-364A-8

```

```

Query Match 20.8%; Score 592; DB 4; Length 3638;
Best Local Similarity 57.3%; Pred. No. 7.8e-120; Indels 90; Gaps 11;
Matches 1335; Conservative 0; Mismatches 905;
QY 157 CTCATTTTTCAGATCACAGCATTTTCAGGAGCACTTTTACCTACACCTGACCTGACCCGGATGCT 216
Db 440 CTCGCTTCCACCTGCTCCGCTTCCGCCAGGCTTTCGCTCGCGCTGGCCCTGACGCC 499
QY 217 CAGTTCCTTGGCTCCGCTTCTCCACTCAGCATCTGGCGCTGCCCTCCAGGGCTACCC 276
Db 500 AGCTTCTTGGCGCGGAATTAAGATCGAGCGCTTCGGGGCTCGAGCGCGCG---GCC 556
QY 277 GGGGCTCTTTCAGACTCGGACGCTGCTTCTTCTTGGGGAGCTGAACCGCGAGCGGAC 336
Db 557 GGGGGGAGCGCGGACTGCTGGCTGCTTCTTCTTGGCACAGTGAATGAGAACGGGAG 616
QY 337 TCGTTCGCTGTGTGAGCCTGTGGGGGGCTTCCCGGAGCCTTTGGCTACCGAGCGCC 396
Db 617 TCGCTGGCGGATGAGCTGTGTCGCGGGCTGGAGCGCTGCTTCTTCTGGCAGGGCAG 676
QY 397 GAGTATGTCATTAGCCCGCTGCCCAATGCTAGCGCGCGCGGCGCAGCGCAACAGCCAG 456
Db 677 GAGTTCACCATCCAGCACAGCGGCTGGGACTCCTTGGACCAGCTTATCGCTGCGAG 736
QY 457 GCGGCACACCTTCTCCAGCGCGGGGTGTTCGGGGCGGGCTTCCGGAGA---CCCCAC 512
Db 737 CGCTGGGGCGCGGACAGCGCGCGAAGACCCCGGGCTCGCTGCGCGGAAAGTTTTCGCC 796
QY 513 CTCCTGCTCGGGGTGGCCTCGGGGTGGAACCCCGCCATCTTACGGGCGCTGGACCTTA 572
Db 797 CTCCTCTCAAGGACTGAGGTGGGAGTGGTAAATGGGCAGGAGAGAGAAAGT 856
QY 573 CAGCGCGCGCGGGGCTTCGGGAGAGTGTAGCGCGCGCAGCTCTGGG--- 624
Db 857 GACACGAAGAGACAAAGACAGGAGGGGTGCTCAAAGAGACAGAACTCC 916
QY 625 -----CGCGCCAAAGCGTTTCTGTGCTTATTCGG 651
Db 917 CGCAAGTGCACACCTTTCGGATCCAAACTAGAACGAAGAGTTTGTGTCGAGGCT 976
QY 652 CGGTAGCTGGAGACGCTGTGTGTCGGGACGAGTCAATGGTCAAGTTCACCGCGCGGAC 711
Db 977 CGCTTCTGGAACACTTCTGTGTGTGATGCTGATGCTGCTTCTATGGGACCGAC 1036
QY 712 CTGGAACATTATCTGCTGAGCTGTGGAACGGCGCGGCTTACCGCATCCCGAC 771
Db 1037 CTGCAACCAATCTCTCAGGTGATGTCATGGGACCGCGAATCTACAAGCACCGGAG 1096
QY 772 ATCCTCAACCCCATCAACATCGTTGTGGTCAAGGTGCTGCTTCTTAGAGATCGTACTCC 831

```

Db 1097 ATCAGGAACCTCCCTCAACCTTGTGGTGAAGTCTAATAGTGAAAGAAAGATGG 1156
QY 832 GGGCCCAAGTACACCGCAATGGGCGCTTGACGCTCGGCACTTCTGTGCTGGGAGAG 891
Db 1157 GGGCCGGAAGTGCAGACAAAGGGGGCTCACACTGCGCACTTCTGACGTGGCAACGG 1216
QY 892 AAGCTGAACAAGTGAAGTGAAGCAACCGGAGTACTGGAGCACTGCCATCTCTTCACG 951
Db 1217 CGTTTCAACAAGCCAGTACCGCCACCGGAGCACTATGACACTGCCATCTGTTCACG 1276
QY 952 AGCGAGACCTGTGTGG- - -AGCCACCACTGTGACACCTCGGCACTGCTGATGGGT 1008
Db 1277 AGACAGAAGTCTGTGGGAAGGAGAGCAGTGTGACACCTCGGGATGSCAGAGTGGC 1336
QY 1009 ACCATGTGACCCCAAGAGAGCTCTCTGTTCATTGAGAGAGTGGGTTTCCATCAGCC 1068
Db 1337 ACCATCTGACCCCAAGAGAGCTCTCAGTGATCAAGGATGAGGAGCTGAGGAGCCG 1396
QY 1069 TTCACCACTGCCACAGAGCTGGGCCACGTGTTCAACATGCCCATGACAACTGTGAAGTC 1128
Db 1397 TACACCTGGCCCATGAGCTAGGGACGTTCTCAGCATGCCCATGATGATTTAAGCC 1456
QY 1129 TGTGAGAGGTGTTGGAGAGTCCGAGCCAAACACATGATTCGCCGACCCCTCATCCAG 1188
Db 1457 TGTGTGAGATGTTGGGCGCATGGGCAAGTACCACATGATGGCGCCATCTTTCATCCAC 1516
QY 1189 ATCGACCGTGCACAAACCCCTGTCAGCTGACGCTGCTGCTGATCATCAGGACATTCCTGAC 1248
Db 1517 GTGAACAAGAGCTGCGCTGGTCTCCTCGAGTGTGTCTACCTCAGAGTCTCTGGAT 1576
QY 1249 AGCGGACGGTGAAGTCTCTCTGACCAACCCAGCAAGCCCATCTCCCTGCGCCGAGAT 1308
Db 1577 GATGGTCAAGGAGATGTTCTTGATGCCCCCACTCGGTTCTGCCCTCCACAGCC 1636
QY 1309 CTGCGGGGCGCCAG- - -CTACACCTTGAGCCAGCAGTGGAGGTGCTTTTGGCGTG 1362
Db 1637 CTCCCGGGCCACAGCAGCACCCTCTACGAGCTGGACAGCAGTGAAGCAGATCTTTGGGCT 1696
QY 1363 GGCTCCAAGCCCTGTCTTTACA-TGCAGTACTGCACCAAGCTGTGTGCACCGGGAAGCC 1421
Db 1697 GATTTCCGACACTGCCCAACACCTCTGAGGAGACATGTGTCCAGCTGTGTCGCCGT 1756
QY 1422 CAAGGACAGATGGTGTGCC- - -AGACCCGCACTTCCCTGGGCGCAT 1467
Db 1757 CATCGGATAGTATGAGCCCAATTTGCCACAAAGATGTAGCTGTCTGGGCTGAT 1816
QY 1468 GGCACAGCTGTGGGAGGCAAGCTCTCCCTCAAGGGGCTCGTGGAGAGACACAC 1527
Db 1817 GGTACACCTGTGGGCGCTGGGACCTGTGCTGGATGGTGTGTACTCAAGGAGAT 1876
QY 1528 CTCACAAAGCACA- - -GGGTGATGTTCTCTGGGCCAAATGGATCCCTATGGCCCC 1581
Db 1877 GTGGAAATCCCAAGCTGTGTAGATGAGACTGGGTCCTTGAGACCTTGGGACAA 1936
QY 1582 TGTCCGCAATGTGTGGGGGCTGTGACGTGGCCAGAGGCACTGACCAACCCACC 1641
Db 1937 TGTCTCGCACCTGTGTGGAGGATACAAATCTCGAACCGTGAATGTGATAATCCAATG 1996
QY 1642 CTGCGCAACGGGGCAAGTACTGCGAGGAGTGAAGTGAATACCGATCTCTGCAACCTG 1701
Db 1997 CTTCAAGTGGAGAGATTTTCTGTTGTAAGAGTCAAGTACCAATCATGCAACACA 2056
QY 1702 GAGCCCTGCCAGCTAGCCCTCCGAAAGAGCTTCCGGGAGGACAGTGTGAGGCTTTC 1761
Db 2057 GAGGAATGTC- - -CACCACGAAAGAAAGCTTCCGGGAGCAGCTGTGAGAAATAT 2110
QY 1762 AAGGGTACACCAACAGACCAACCGGCTACTCTCGCCGTGGATGGGTGCCCAAGTAC 1821
Db 2111 AATGCTTACACCA- - -CACTGACCTGGATGGGAATTTCTCTGAGTGGGTCCCCAAGTAT 2167
QY 1822 TCCGGGCTGTCTCCCGGGACAAAGTGCAGCTCATCTGCCGAGCAATGGCACTGGCTAC 1881

Db 2168 TCAGAGTGTCCCCCGAGACCGATGCAAGCTGTTTTGCAGAGCCCGTGGAGAGTGA 2227
QY 1882 TTCATATGTGTGGCACCCCAAGGTGGTGGACCGCAGCTGTCTCTCTGACATCCACCTCC 1941
Db 2228 TTCAAAGTGTTTGAAGCTTAAGGTGATCGATGGCACTCTGTGTGACGGGATCTCTGTCC 2287
QY 1942 GTCTGTGTCCAAAGCAAGTGCATCAAGGCTGGCTGTGATGGGAACCTGGGCTCCAAAG 2001
Db 2288 ATCTGCTCCGGGGCAATGTGTTAAGGCTGGCTGTGACCATGTGGTGAACCTACCTAAG 2347
QY 2002 AGATTCCGAAGTGTGGGGTGTGGGGGAGACAATAAGAGCTGCAAGAAGCTGACTGGA 2061
Db 2348 AAGCTGGACAAATGTGGGGTGTGGGGCAAGGCACTGCTCTAGGAAGATCTCCGGT 2407
QY 2062 CTCTTCCACCAAGCCATGCATGGCTACAATTTCTGTGTGGCCATCCCGCAGGCGCTCA 2121
Db 2408 TCTTTACCCCTTTCAGTTATGGCTACAAATGACATTTGACCATCCAGCTGGTGGCACA 2467
QY 2122 AGCATCGACATCCCGAGCGGTTTACAAGGCTGATCGGGGATGACAACTACCTGGCT 2181
Db 2468 AACATGTGTGAACACGCGGAGTCAACGAGGTGAGGAACGAGCGAGCTACTGGCG 2527
QY 2182 CTGAAGAACACCCAAAGCAAGTACTGCTCAACGGGCAATTTCTGTGTGGCGGTGGAG 2241
Db 2528 CTGAAGACAGCAATGGGCACTGCTCAATGTTAACTGGGCACTCTCTGCCATAGAG 2587
QY 2242 CGGACCTGGTGTGAAGGCGAGTCTGCTGGGTTACAGCGGCAAGGCGGACAGCGGTGGAG 2301
Db 2588 CAAGACATCTTGGTGAAGGGGACCATCTGAAATGACAGTGGCTCCATGGCTACCGTGGAG 2647
QY 2302 AGCTGTGAGGCTTCCCGGCCCATCTGAGCGCTGAGCGGTGGAGTCTCTCCGTTG- - - 2358
Db 2648 CGGCTGCAGAGCTTCCAGGCCCTGCTGAGGCTCTTACAGTACAGCTCTCTGACTGTCT 2707
QY 2359 GGGAGATGACACCGCCCGGTCGGTACTCTCTTCTATCTGCCCCAAGA 2408
Db 2708 GGTGAGGTCTTCCCTCCAAAAGTCAGATATACCTTCTTTGTCCCAATGA 2757

RESULT 6
US-09-122-126B-1
; Sequence 1, Application us/09122126B
; Patent No. 6451575
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: AGGRECAN DEGRADING METALLO PROTEASES
; FILE REFERENCE: DM6909
; CURRENT APPLICATION NUMBER: US/09/122,126B
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 4192
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (406)..(2916)
US-09-122-126B-1

Query Match 20.5%; Score 584.8; DB 4; Length 4192;
Best Local Similarity 60.4%; Pred. No. 3e-118;
Matches 1094; Conservative 0; Mismatches 692; Indels 24; Gaps 7;

QY 609 CCGCGCAGGTCTGGGCGGCCCAAGCGTTTCTGTCTATCCCGGTTACGTGGAGAGCT 668
Db 1014 CAGCCCCAGACCCCGAAGAGCAAGCGCTTGTCTTCACTAGTAGATTTTGTGGAGACT 1073
QY 669 GGTGCTCGCGACAGTCAATGGTCAAGTTTCCACGGCGGAGCTGGAACTATCTGCT 728
Db 1074 GGTGCTGGCAGATGCAAGATGGCCGATTCACGGTGGGGCTTAAGGCTTACTGCT 1133
QY 729 GAGCTGCTGGCAACGGCGGCGAGCTCTACCGCCATCCCGCATCTCAACCCCATCAA 788

Db	1134	AA	CAAGTGA	TGGCAGCAGCAGCCAAAGCCCTTCAAGCACCAAGCATCCGAATCCTGTCAAG	1193
Qy	789	CATCGT	TGTGGTCAAGGTGCTCTTCTTAGAGATCGTGACTCCGGGCCCAAGGTACACGG	848	
Db	1194	CTTGGT	TGGTGA	CTCGCTAGTGATCCTGGGTGTCAGGCAGGAGGGGCCCAAGTGGGGCC	1253
Qy	849	CAATGGG	CCCTGAGCGTGC	GCACACTCTGTGCTGTGGCAGAGAAGCTGAACAAATGTAG	908
Db	1254	CAGTGT	CGCCAGACCC	TGCGCAGCTTCTGTGCTGTGGCAGCGGGCCCTCAACACCCCTGA	1313
Qy	909	TGACAAG	CACCCGAGTACTGGGACACTGGACACTGCCATCCTCTTACACAGGCAGACCTGTGTGG	968	
Db	1314	GGACT	CGGACCTTGACACTTTGACACAGCCATCTGTGTACCCGTCAGGACCTGTGTGG	1373	
Qy	969	AGCCAC	CACTGTGACACCCCTGGGCATGGCTGATGTGGGTACCATGTGTGACCCCAAGAG	1028	
Db	1374	AGTCTC	CACTTGGACACGCTGGGTATGSGTGTGATGTGGGCACCGCTGTGACCCGGCTCG	1433	
Qy	1029	AAC	TGCTGTGTCATTTAGGACGATGGCTTTCATCAGCCTTACCACTGCCACAGACT	1088	
Db	1434	GAGCT	GTGCCATTTGGAGGATGATGGGCTCCAGTCAAGCTTCACTGCTGCTCATGA	1493	
Qy	1089	GGCC	CACGTGTTCAACATCCCCCATGACAATGTAAATCTCTGACAGAGTGTGTGGAA	1148	
Db	1494	GGGT	CATGCTTCAAGATGCTCCATGACAACATCCAAGCCATGCATCAGTTTGAATGGCC	1553	
Qy	1149	GCTC	GAGCCAA----CCATGATGTCCCGAGCCCTCATCCAGATCGACCGTGCCACACC	1205	
Db	1554	TTT	GACACCTCTCGCCATGTATGGCCCTCTGTATGSGCTCATGTGATCCTGAGGAGCC	1613	
Qy	1206	CTGGT	CAGCTGCAGTGTGCCATCATCACGACTTCTCTGACAGCGGCACGCTGACTG	1265	
Db	1614	CTGGT	CCCCCTGCAGTGC	CCCGCTTCATCACTGACTTCTCTGACAAATGGCTATGGGCACTG	1673
Qy	1266	CTCT	CTGGACCAACCCAGACGCCCATCTCCCTGCTCCGAGAGATCTGCCGGCGCCAGCTA	1325	
Db	1674	TC	TCTTAGACAACACAGAGCTCCATTCGATCTGCTGTGACTTCCCTGGCAAGACTA	1733	
Qy	1326	CAC	CTTGACCCAGTGGAGCTGGCTGCTTTGGCTGGGCTCCAGCCCTGTCTCTTACAT	1385	
Db	1734	TGAT	GTGACCGCAGTGCCAGCTGACCTTGGCGCCGACTCACGCCATGTCTCCACAGCT	1793	
Qy	1386	---	GCAGTACTGCACCAAGCTGTGTGACCGGGAAGCCAAAGGACAGATGTGTGCCA	1442	
Db	1794	GC	CGCCCTGTGTGCTCTGTGTGCTGTGGCCACCTCAATGGCCATGCCATGTGCCA	1853	
Qy	1443	GAC	CGCGCATTCCTCTGGCCGATGGCACACAGCTGTGGCAGGGCAGCTTGCCTCAA	1502	
Db	1854	GAG	CAAACTGCCCCCTGGCCGATGGCACACCTTGGCGGCCCCACAGGCCTGCATGGG	1913	
Qy	1503	AGGG	CGCTGCTGGAGAGACAACTCTCAACAAAGCACAGGCT-----GGATGGTTCTCTG	1556	
Db	1914	TGGT	CGCTCCACATGGACCTCCAGGACTTCAATATCCACAGGCTGTGTGGCTG	1973	
Qy	1557	GG	CAATGGGATCCCTATGGCCCTGTCTGGGCAACATGTGGTGGGGCGTGCAGCTGGC	1616	
Db	1974	GGT	TCTTGGGACCACTGGGGTGA	CTCTCGGACCTTCTGGGGGTGGTGTCCAGTCTCTC	2033
Qy	1617	CAG	GAGCAGTGCACCAACCCACCCCTGCCAACGGGGCAAGTACTGCCAGAGAGTGAG	1676	
Db	2034	CT	CCGAGACTGCACGAGCCTGTCCCCCGGAATGGTGGCAAGTACTGTGAGGGCCCGC	2093	
Qy	1677	GGT	GAATACCGATTCCTGAACTGGAGCCCTTGCCCCAGCTCAGCTCCCGGAAGAGCTT	1736	
Db	2094	TAC	CGCTTCCGCTCTGCAACACTGAGACTTGCCCAA---CTGGCTCAGCCCTGACCTT	2150	
Qy	1737	CCGG	GAGGACGTGTGAGGCTTTCAAGGCTTACAACACAGCAGCAACCGGCTCACTCT	1796	
Db	2151	CCG	GAGGACGTGTGCTGCCTTACAACCGCACCGACCTCTTCAAGAGCTTCCAGG	2210	
Qy	1797	CG	CGTGGCATGGGTGCCCAAGTACTCGCGGCTGTCTCCCCGGGACAAGTGCAGCTCAT	1856	

```

RESULT 7
US-09-122-126B-14
; Sequence 14, Application US/09122126B
; Patent NO. 6451575
;
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: AGGREGAN DEGRADING METALLO PROTEASES
; FILE REFERENCE: DM6909
; CURRENT APPLICATION NUMBER: US/09/122,126B
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 3250
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (121)..(2910)
US-09-122-126B-14

```

	Query Match	19.2%	Score 547.4	DB 4	Length 3250
	Best Local Similarity	57.7%	Pred. No. 3.8e-110		
	Matches 1043	Conservative	0	Mismatches 746	Indels 18
	Gaps				
Qy	617	GGTGTGGCGCGCAACGGCTTTTCGTGTCTATCCCGCGGTACGTGGAGACGCTGGTGGTCTG	676		
Db	884	GGTGGCGCGCGCGCTCCATCTCCGGCGCGCGCCAGGTGGAGCTGCTTCTGGTG	943		
Qy	677	CGGACGAGTCAATGGTCAAGTCTCCAGCGCGCGACCTGGAACATTTATCTGCTGACGCTGC	736		

Db	944	CTGAGCGCTCATCGCGCGGTGTATGGCCGGCGGCTCGAGCATTAACCTGTGACCCCTGG	1003
QY	737	TGGCAAGGGGGCGCGACTCTACCGGCATCCCGCACTCTCAACCCCACTCAACATCGCTTG	796
Db	1004	CCTCCATCGCCAATAGGCTGTACAGCCATGCTAGCATCGAGAACCACATCCGCCTGGCCG	1063
QY	797	TGGTCAAGGTGCTGCTTCTTAGAGATCGTGACTCGGGGCCCAAGGTCAACGGCAATCGG	856
Db	1064	TGGTGAAGGTGGTGGTCTTAGCGCAAGAGCAAGAGCCGTGAAGTGAAGCAAGACGGTG	1123
QY	857	CCCTGAGCGCTGCGCAACTTCTGTGCGCTGGCAGAAGAGCTGAACAAAGTGAAGTGCACAGC	916
Db	1124	CCACCACACTCAAGAACTTTGCAAGTGGCAGCACCAACACAAACAGCTGGGAGATGACC	1183
QY	917	ACCCCGAGTACTGGGACACTGCCATCCTCTTACCAGGCGAGGACCTGTGTGGAGCCACCA	976
Db	1184	ATGAGGAGCACTAGATGACGCTATCCTGTTACTCGGGAGGATTATGTGGGCATCATT	1243
QY	977	CCTGTGACACCTGGGCGATGGCTGTATGGGTACCATGTGTGACCCCAAGAGAAGCTGCT	1036
Db	1244	CATGTGACACCTGGGAATGGCAGAGCTTGGGACCATATGTTCTCGAGCGCAGCTGTG	1303
QY	1037	CTGTCAITTAGGAGGATGGCTTCCATCAGCCTTACCACCTGCCACAGCTGGGCCACG	1096
Db	1304	CTGTGATTGAAGAGATGGCTCCAGCGAGCCTTCACTGTGGCTCACGAAATCGGACATT	1363
QY	1097	TGTTCAACATGCCCCATGACAATGTGAAAGTCTGTGAGGAGGTGTTGGGAAGCTCCGAG	1156
Db	1364	TACTTGGCCTCTCCCATGAGCATCCAAATCTGTGAAGAGACCTTGGTTCACACAG	1423
QY	1157	CCAACCACATGATGTCCCGACCCCTCATCCAGATCGACCGTGCACCCCTGGTCAAGCCT	1216
Db	1424	ATAAGCCTTAAATGTCTTCCATCTTACACAGATTTGATGCATCTAAGCCCTGGTCCAAAT	1483
QY	1217	GCAGTGTGCGCATCATCACCGACTTCTCGACAGCGGCACGGTGACTGCTCTCTGACCC	1276
Db	1484	GCATTGAGCCACCATCACAGAAATCCTGGATGTGGCCATGGTAACHTGTTTGTGTGACC	1543
QY	1277	AACCCAGCAAGCCCATCTCCCTGCCGAGGATCTGCCGGGGCGGAGCTTACACCTTGAGCC	1336
Db	1544	TACCACGAAAGCAGATCCTGGGCCCCGAGAAGTCCAGGACAGACCTACGATGCCACC	1603
QY	1337	AGCAGTCGAGCTGGCTTTTGGGTGGGCTCCAAACCCCTGTCTTACATGCAAGTACTGCA	1396
Db	1604	AGCAGTCAACCTTGACATTCGGGCGCTGAGTACTCCGCTGTGCCGGCATGGATGCTGTG	1663
QY	1397	CCAAGCTGTGGTCACCGGGAAGGCCCAAGGACAGATGCTGCCAGACCCGCCACTTCC	1456
Db	1664	CTGCGCTGTGGTGTGTGTGTACGCCAGGGCCAGATGGTCTGTCTGACCAAGAAGCTGC	1723
QY	1457	CCTGGGCGCATGACACAGCTGTGGCGAGGGCAAGCTCTGCCTCAAAAGGGCGCTCGCTGG	1516
Db	1724	CTGCGGTGGAAGGACGCGCTTGTGGAAGGGGAGGAATCTGCCTCGAGGGCAATGCTGG	1783
QY	1517	AGAGACACAACCTCAACAAGCAGGGTG-----GATGTTCTCTGGGCCAAATFGGATC	1570
Db	1784	ACAAAACACAGAAAAAATATTATCAAGCTCAAGCATGGCAACTGGGGATCTTGGGGAT	1843
QY	1571	CCTATGGCCCTCTCGCGCACATGTGTGGGGCGTCCAGCTGGCCAGAGGCGCATGCA	1630
Db	1844	CCTGGGGCCAGTGTCTCGCTCATGTGAGGAGGAGTGCAGTTTGGCTATGTCGTAAGTGA	1903
QY	1631	CCAACCCCAACCCCTGCGCAACGGGGCAAGTACTCGAGGGGAGTCAGGTTGAATATCCGAT	1690
Db	1904	ATAACCTGTCTCCAGAAACACGAGCAGCTACTGCACAGGAAAGAGGCCACTTACCGCT	1963
QY	1691	CCTGCAACCTGGAGCCCTGCCCGAGCTCAGCCTCGGAAAGAGCTTCCGGGAGGAGCAGT	1750
Db	1964	CCTGCAAGTCTCATGCGCCTGCC-----CACCCAATGGTAAATCATTTTCGTCATGAACAGT	2017
QY	1751	GTGAGCTTTCAACGCTTACACACAGCACCAGCCGCTCACTCTCGCGGTGGCATGGG	1810

```

RESULT 8
US-09-369-364A-1
; Sequence 1, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3002
; TYPE: DNA
; ORGANISM: mus musculus ADAMTS-5.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (18)..(2810)
US-09-369-364A-1

```

Query Match	19.1%	Score 545.8;	DB 4;	Length 3002;
Best Local Similarity	57.7%	Pred. NO. 8.2e-110;		
Matches 1042; Conservative	0;	Mismatches 747;	Indels 18;	Gaps 3;

FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(5648)
; NAME/KEY: misc_feature
; LOCATION: (1406)
; OTHER INFORMATION: n=T
; NAME/KEY: misc_feature
; LOCATION: (1563)
; OTHER INFORMATION: n=T
US-09-369-364A-12

Query Match 14.4%; Score 411.2; DB 4; Length 5804;
Best Local Similarity 54.4%; Pred. No. 1.7e-80;
Matches 971; Conservative 0; Mismatches 785; Indels 30; Gaps 6;

QY 629 CCAAGCGTTTCGTCTATCCCGGTACGTGGAGAGCTGTGGTGGCGGAGAGTCAA 688
DB 700 CAAAGCTTTTATCCCTATCCAGCGTTTGTAGAAGTCTTGGTGGCGGAGACAGAA 759
QY 689 TGGTCAAGTCCAGCGCGGAGCTGGAACATTATCTGTGAGCTCTGGCAACGGCG 748
DB 760 TGGTTTCACTACCATGGAGAAACCTTCAACACTATATTTTAACTTTAATGTCAATTGTAG 819
QY 749 CGGACTCTACCGGCATCCAGCATCTCAACCCATCAACCTGTTGTGGTCAAGTGC 808
DB 820 CCTCTATCTATAAAGACCCAGTATTGGAATTTAATTAATTTAATTTAATTTAATTTAATTTA 879
QY 809 TGTCTTCTAGAGATCGTGACTCCGGGCGCAAGGTCAACGGCATCGGGCCCTGACGCTGC 868
DB 880 TTGTGATTCATTAAGACAGATGGGCTTCCATATCTTTTAACTGCTACACACATTA 939
QY 869 GCAACTTTGTCTGCTGGCAGAGAGTGAACAAAGTGAAGTGAAGCAAGCCCGAGTACT 928
DB 940 AAAACTTTTGGCAGTCCAGCATTCGAACAGTCCAGGTGAATCCATCA 988
QY 929 GGGACATGCCATCCTCTTACAGGAGGAGGAGCTGTGGAGC---CACACCTGTGACA 985
DB 989 -TGATACTGCTGCTCTTCAACAGAGAGGATATCTCAGAGCTTCACGACAAATGTGATA 1047
QY 986 CCCTGGGCGATGGCTGATGTGGGTACCACTGTGTGACCCCAAGAGAGTGTCTGTCTAATG 1045
DB 1048 CTTTAGGCTGGCTGAAGTGGGAAACCATTTGTGATCCTATAGAAGCTGTCTATTAGT 1107
QY 1046 AGGAGATGGCTTCCATCAGCTTACCAGTCCAGGAGCTGGGCGAGCTGTTCACA 1105
DB 1108 AAGATAGTGATTTAGTACAGCTTTTACGATCGCCCATGAGCTGGGCGATGTGTTAA 1167
QY 1106 TGCCCATGACAATGTGAAGTCTGTGAGAGGTGTGTGGAGCTCCGAGCCCAACACA 1165
DB 1168 TGCCTCATGTGACACACAAATGTAAAGAGAGGAGTTAAG---AGTCCCGAGCATG 1224
QY 1166 TGATGTCCCGACCTCATCCAGATTCGACCTGCCACCCCTGTGAGCTGCGAGTCTG 1225
DB 1225 TCATGGCTCCAACAGTGAATTTACACCAACCCCTGGATGTGTGCAAGTGTAGTCGAA 1284
QY 1226 CCATCATCAGGCTTCTGTGACAGCGGAGCTGCTGCTCTGACCAACCC---CA 1282
DB 1285 ATATATCAGTGTGTTTGAACATCTGTTATGCGAGTGTGCTTTAACAACCTGAAT 1344
QY 1283 GCAAGCCCATCTCCCTGCCAGGATCTGCCGGGCGGAGCTACACCTGAGCCAGCAGT 1342
DB 1345 CCAGACCTTACCTTTGGCTGTCCAACTGCCAGGATCCTTTTCAACAGTGAATAAACAAT 1404
QY 1343 GCGAGCTGGCTTTTGGCGTGGCTCCAGCCCTGTCTTACATGACGTACTGACCAAGC 1402
DB 1405 GNGAATGATTTTGGACAGGCTTCTCAGGTGTGCCATATATGATGACGTGACAGCGC 1464
QY 1403 TGTGTGACCGGGAAGCCCAAGGACAGATGTTGTGCGACCGCCGACCTTCCCTGGG 1462
DB 1465 TCTGTGGAATAACCTCAATGGAGTACAAAGGCTGCCGAGCTAGCACACACCTGGG 1524
QY 1463 CGATGGGACAGCTGTGGCGAGGGAAGCTCTGCTCTCAAGGGGCTGCGTGGAGAGAC 1522

DB 1525 CGATGGGACGAGTGGCGCTCGAGAGCTGCAAGNATGGATTTTGTGTTCCCAAG 1584
QY 1523 ACAACCTCAACAAGCACAGGTTGATGTTCTCTGGGCCAAATGGGATCCCTATGGCCCT 1582
DB 1585 AAA---TGGATGTCCCGTGCAGATGATCCTGGGGAAGTTGGAGTCCCTTTGGAACCT 1641
QY 1583 GCTCGCACATGTGTGGGGCGTGCAGCTGGCCAGGAGCAGTGCACCAACCCACCC 1642
DB 1642 GCTCCAGAACATGTGGAGGGGCATCAAAACAGCCATTCGAGAGTGCACAGACAGAAC 1701
QY 1643 CTGCCAACGGGGCAAGTACTCGAGGAGTGAAGGTGAAATACCGATCTCTGCAACCTGG 1702
DB 1702 CAAAATATGGTGAATAATCTGTGTAGACGTAGATGAATTTAAGTCTCTGCAACGG 1761
QY 1703 AGCCCTGCCCGAGCTCAGCCTCCGGAAGAGCTTCCGGAGGAGCAGTGTGAGGCTTCA 1762
DB 1762 AGCCATGTCTCAAGCAGAAGCGAG-----ACTTCCGAGATGAACAGTGTCTCCTACCTTTG 1815
QY 1763 AGGGCTACACACAGCACCACCGCTCACTCTCGCGTGCATGGCTGCCCAAGTACT 1822
DB 1816 ACGGGAAGCATTTTAAACATCAACGCTCTGCTTCCCAATGTGGCTGGGTCTCTTAATACA 1875
QY 1823 CGGGCTGTCTCCCGGGACAAAGTCAAGCTCATCTGCCGAGCCAAATGGCACTGSCAT 1882
DB 1876 GTGGAATCTGTATGAAGACCGGTGCAAGTTGTTCTGCAGAGTGGCAGGAACA-AGCCT 1935
QY 1883 TCTATGTGCTGCACCAAGGTGTGGGAGAGACAATGAAGCTGCAAGAGTGAAGTGGAC 2062
DB 2056 GGAGAGATAAATCGGGGTTTGTGGTGGCGATAATTTCTTCAATGCAAAACAGTGGCAGGA 2115
QY 2063 TCTTCAACAGCCATGATGCTACAATTTGTTGGTGGCCATCCCGCAGCGCTCAA 2122
DB 2116 CATTTAATACAGTACATATATGTTTCAATACGTGCTCGGAATTCAGCTGTGTACCA 2175
QY 2123 GCATCAGATCCGCCAGCGGCTTACAAAGGCTGATCGGGATGACAACTACCTGGGCTC 2182
DB 2176 ATATTGATGTGGCAGCACAGTTTCTCAGGGGAAACAGACGATGACAACTACTTAGCTT 2235
QY 2183 TGAAGAACAGCAGCAAGTACCTGCTCAAGGGCATTTCTGTTGCTGGCGGTGGAGC 2242
DB 2236 TATCAAGCAGTAAGGTGAATTTCTGCTAAATGGAAACTTTGTGTCAATGGCCAAA 2295
QY 2243 GGGACCTGTGTGAAGGCGAGTCTGCTGGGTACAGCGGACAGCGGAGCGGTGGAGA 2302
DB 2296 GGGAAATTCGATTTGGAAATGCTGTGTAGTAGTACAGTGGTCCGAGACTGCCGTAGAAA 2355
QY 2303 GCCTGACGCTTCCCGGCCCATCTCTGGAGCGGTGACCGTGGAGGCTCTTCCGTGGGGA 2362
DB 2356 GAATTAATCAACAGATCGCATTTAGCAAGAACTTTTGTCTCAGGTTTGTGCGTGGGAA 2415
QY 2363 AGATGACACCGCCCGGCTCCGCTACTCTCTATCTGCCCCAAAGA 2408
DB 2416 AGTTGTACACCCCGATGACGCTATTCTTTCAATATTCATTTGA 2461

RESULT 11

US-09-130-491-7
; Sequence 7, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491

QY	1204	CCCTGGTACGCTGCACTGCTGCCATCATCAACCGACTTCTCTGGACAGCGGCACGGTGAC	1263
DB	21		
QY	1264	TGCTCTCTGGACCAACCCAGCAAGCCCATCTCCCTGCCGCCGAGGATCTGCCGGGCGCCAGC	1323
DB	81		
QY	1324	TACACCTTGAGCCAGCAGTGCAGAGTGGCTTTTGGCGTGGGCTCCAAAGCCCTGTCCTTAC	1383
DB	130	-----GCAAGCAGATCTTTGGGCTGATTTCCGACACTGCCCCACACCTCTGTG	179
QY	1384	ATCGCACTACTCACCAAGCTGTGTGCACCGGGAAGGCCAAGGACACAGATGTGTGC--	1441
DB	180		
QY	1442	----AGACCGGCACCTTCCCTGGGCCGATGSCACCACTCTGCGGAGGCGCAAGCTGCG	1497
DB	240		
QY	1498	CTCAAGGGGCTCGGTGGAGAGACACAACCTCAACAGCACAC-----GGGTGGATGGT	1551
DB	300		
QY	1552	TCTGCGGCCAAATGGATGCCCTATGGCCCTCTCGGCCACATGTGTGGGGGGGTGCGAG	1611
DB	360		
QY	1612	CTGGCCAGGAGCAGTGCACCAACCCACCCCTGCCAACGGGGGCAAGTACTGCGAGGGA	1671
DB	420		
QY	1672	GTGAGGGTGAATACCGATCCTGCNAACCTGGAGCCCTGCCGCCAGCTCAGCCTCCGGAAAG	1731
DB	480		
QY	1732	AGCTTCCGGGAGGACAGTGTGAGCTTTTCAACGGCTACAACACACCAACCCCGGCTC	1791
DB	534	AGCTTCAGGACGACGAGTGTGAATAATCCAGCACCTCAGAATGGAGAAAGATTTGCCTGGGA	1791
QY	1792	ACTCTGCCGTGGCATGGGTGCCCAAGTACTCCGGCGTGTCTCCCGGGGACCAAGTGCGAC	1851
DB	591	GGGAATTTCTTCACTGGGGTCCCCAAATACTCAGGAGTGTCCCCCGGAGACCGATGC	650
QY	1852	CTCATCTCGCAGGCAATGGCACTGGCTACTTCTATGTCTGGCACCCCAAGTGTGTGGAC	1911
DB	651		
QY	1912	GGCAGCTGTGCTCTCTGACTCCACCTCTCGTCTGTCTCCAAGCAAGTGCATCAAGGCT	1971
DB	711		
QY	1972	GGCTGTGATGGAACTCTGGGCTCCAGAAGAGATTCGACAAGTGTGGGGTGTGTGGGGA	2031

Db 235 TGATCGTAGAAGATGAAATATGGGGCCAGAGGTGTCGACATGGGGGCTTACACTGC 294
 Qy 869 GCAACTTCTGTGCTGGCAGAGAGCTGAACAAGTGAAGTGAACAAGCAGCCCGAGTACT 928
 Db 295 GTAACTTCTGCACTGGCAGCGGCTTTCAACAGCCGACCGCCACCCAGACACT 354
 Qy 929 GGGACACTGGCATCTCTTCCACGAGGAGACCTGTGTGG- --AGCCACACACTGTGACA 985
 Db 355 ACGACAGCGCATCTCTCTCACCAGACAGAACTTCTGTGGGACAGAGGGGCTGTGTGACA 414
 Qy 986 CCCTGGGCATGGTGTGATGGGTACCATGTGTACCCCAAGAGAGTGTCTGTCTATTG 1045
 Db 415 CCCTGGGTGTGGCAGACATCGGACCACTTTGTGACCCCAACAAAGCTGCTCGGTATCG 474
 Qy 1046 AGGACGATGGCTTCCATCAGCTTCCACCACTGCCACAGAGCTGGGCCAGCTGTCAACA 1105
 Db 475 AGGATGAGGGCTCCAGCGGCCACACACCTTGGCCATGAACTAGGCGAGTCTCTAGCA 534
 Qy 1106 TGCCCCATGACAATGTGAAGTGTGTGAGGAGGTGTTTGGGAAGTCCCGAGCCAAACACA 1165
 Db 535 TGCCCCAGCAGACTTCCAAGCCCTGACACAGGCTTCTCGGGCCCATGGGCAAGCACCAG 594
 Qy 1166 TGATGTCCTCCGACCTCATCCAGATCACCGTGCACACCTTGTGAGCTGCAAGTGCAGTCTG 1225
 Db 595 TGATGGACACCTGTTGTCACCTGAACACAGAGCTGGCCCTGGTCCCCCTGACGCGCCA 654
 Qy 1226 CCATCATCACCGACTTCTCTGGACAGCGG 1254
 Db 655 TGTTCAGGGTGCACCTGCAGGGTGG 683

RESULT 13

US-09-392-184-5/c
 ; Sequence 5, Application US/09392184
 ; Patent No. 6395889
 ; GENERAL INFORMATION:
 ; APPLICANT: Robison, Keith E.
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN
 ; TITLE OF INVENTION: PROTEASE HOMOLOGS
 ; FILE REFERENCE: 5800-55
 ; CURRENT FILING DATE: 1999-09-09
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 5
 ; LENGTH: 5357
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(5357)
 ; OTHER INFORMATION: reprotolysin (ADAM family of metalloprotease)
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(5357)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-392-184-5

Query Match 8.7%; Score 247.2; DB 4; Length 5357;
 Best Local Similarity 51.8%; Pred. No. 7.5e-45;
 Matches 927; Conservative 0; Mismatches 778; Indels 83; Gaps 13;

Qy 630 CAAGGTTTCGATATCCCGCGGTAGCTGGAGACGCTGGTGGTGGGAGAGTCAAT 689
 Db 4617 CCAGCGGTGCGTCAAGAAAGAGTGGGTGGAGACCTGGTAGTGTATGATCCAAAT 4558
 Qy 690 GGTCAAGTTCCAGG- --CGCGACCTTGAACATTATCTGTGACGCTGTGCGCAACGGC 746
 Db 4557 GGTGAGTACACGAGAGCGGAGGTGTGAGAGTATGTGTGACCATCATGACATGGT 4498
 Qy 747 GCGCGACTTACCGCCATCCAGCATCTCAACCCCATCAACATCTGTGTGGTCAAGGT 806
 Db 4497 GCGTGGCTGTTTCATGACCCAGCATTTGGGAACCCCATCCATCACCATTGTGCGCT 4438

Qy 807 GCTGCTTCTTAGAGATCGTGAATCGGCGCCCAAGGTCTACCGGCAATGGCGCCCTGACGCT 866
 Db 4437 GGTCTCTGTTGAAGATGAGGAGGAGGACTTAAGATACGACCATGACAGACAACACCT 4378
 Qy 867 GCGCAACTTCTGTGCTGGCAGAGAAAGCTGAACAAGTGAAGTGAACAAGCAGCCGAGTA 926
 Db 4377 GAAGAGCTTCTGAAGTGGCAGAAAAGCATCAACATGAAGGGGATGCCATGCTCTGCA 4318
 Qy 927 CTGGGACACTGTCATCTCTTCCACGAGGACGACCTGTGTGGAGCCACA- - - - -CCTG 980
 Db 4317 CCATGACACTGTCATCTCTGTCACCAAGAAAGACCTGTGTGACGACATGAACCGCCCTG 4258
 Qy 981 TGACACCTGGGCATGGTGTGATGGGTACCATGTGTGACCCCAAGAGAAAGTCTCTGT 1040
 Db 4257 TGAGACCTGGGACTGTCCCATGTGGCGGATGTGCCAGCCGACACGACCTGCAAGCAT 4198
 Qy 1041 CATTGAGGACGATGGCTTCCATCAGCTTCCACCACTGCCACAGCTGGCGCAGCTGT 1100
 Db 4197 CAACGAGACAGCGGCTGCGCTGCTCTACTGTAGCCCAAGAGCTCGGCGACAGTTT 4138
 Qy 1101 CAACATGCCCCATGACAATGTGAAGTGTGTGAGGAGGTGTTTGGGAAGTCTCCGAGCCAA 1160
 Db 4137 TGGCATTCAGCATGACGGAAGCGCAATGACTGTGAGGCCGTTGGGAACACGAC- - - - -C 4084
 Qy 1161 CCACATGATGTCCTCCGACCTCATCCAGATGACCGTGCACACCTGTGTGAGCTGAG 1220
 Db 4083 TTTTCATGATGTCCTCAGAGCTCTGTAGACCCGCTCCCTCAGCTGTGCTGCGTGGAG 4024
 Qy 1221 TGTGTCATCATCACCGACTTCTCTGGACAGCGGACGCTGTGACTGCTCTCTGGACCAAC 1280
 Db 4023 CGGCCAGTATATCACAGGTTCTTTGACCGTGGTGGGCGCTGTGCTTGGACAGCCCTCC 3964
 Qy 1281 CAGCAAGCCC- - -ATCTCCTGCCGAGGATCTGCGGGCGCCAGCTACACCTCAGCA 1337
 Db 3963 TGCACAGGACATATCGACTTCCCTCGGTGCCACTGCGCTCTCTATGATGAAGCA 3904
 Qy 1338 CGAGTGCAGGCTGGCTTTTGGGCTGGGCTCCCAAGCCCTCTCTTACATGACGTA- - - - -CTG 1394
 Db 3903 CAGTGCCTGCTCCAGTACGGGCGCTACTTCTGCTGCGAGGACATGGATAATGCTG 3844
 Qy 1395 CACCAAGCTGTGTGACCGGAAAGCCCAAGGACAGATGTGTGCCAGACCCGCACTT 1454
 Db 3843 CCACACACTCTGTGCTGTGTGGGACCCACTGTCTACTCCAAAGCTGGATGACGCT- - - - - 3788
 Qy 1455 CCCTGGGCGGATGGCAGCAGCTGTGGCAGGCGGAGCTCTGCCCTCAAGGGCGCTGGT 1514
 Db 3787- - - - -GTGGACGCGCCCGGTGTGGGAGAAATAGTGGTGTCTCAGTGGGAGTGGT 3735
 Qy 1515 GGAGAGACACACCTCAACAAGCAGAGGTGGATGGTCTCTGGGCAAAATGAGATCCCTA 1574
 Db 3734 ACCGTGGGCTTCCGCGCCGAGGCC- - -GTGGATGGTGGCTGTGCTGGCTGGAGCGCTG 3678
 Qy 1575 TGCCCTCTGCTGCGCAGATGTGGTGGGCGGTGTGAGCTGGCCAGGAGGACGTGACCA 1634
 Db 3677 GTCCATCTGCTACGGAGCTGTGGCATGGGCGGTACAGAGCGCGGAGCGAGCTGACGCA 3618
 Qy 1635 CCCCACCTTCCCAACGGGGCAAGTACTGCGAGGGAGTGAAGTGAATACCGTCTCTG 1694
 Db 3617 GCCTACGCCCAATACAAAGGAGATACTGTGTGGTGTGAGCGAAGCGCTTCGCGCTCTG 3558
 Qy 1695 CAACCTGGAGCGCTTCCGCGCCAGCTCAGCTCCGGAAGAGCTTCCGGGAGGAGCATGTGA 1754
 Db 3557 CAACCTGACGCTGCTG- - - - -CTGCTGGCGCCCTCTCTCCGCGACGTCCAGTGCAG 3504
 Qy 1755 GGTCTTCAACGGGTACACCAAGCAGCAACCGGCTACTCTGCGCGTGGCATGGGTGCC 1814
 Db 3503 CCACCTTTCAGCTGTCTCTACAAAGGCGCAGCTGCA- - - - -CATGGTGGC 3456
 Qy 1815 CAGTACTCGGCGTGTCTCCCGGGGACAAAGTGAAGTCAAGTCTATCTGCGGAGCAATGGCAC 1874
 Db 3455 CGTGGTCAATGAGCTGAACCCC- - - - -TGCGAGCTGCACTGCGCGCGCGCGNATGA 3405
 Qy 1875 TGGCTACTTCTATGTGTGCGCACCCCAAGGTGTGGAGCGGACGCTGTGTCTCTCTGACTC 1934

```

Db 3404 GTACTTTGCCGAGAAGCTGCGGGACGCGTGGTGCATGGCAACCCCTGCTACCAAGTCCG 3345
QY 1935 CACCTCG-----TCGTGTCCAGCAAGTGCATCAAGCTGGCTGTGATGGGAACCT 1988
Db 3344 AGCCAGCGGGACCTTGCATCAACGGCATCTTAAGAACGTGGGTGTGACTTCGAGAT 3285
QY 1989 GGGCTCCAAGAAGAGATTGCACAAAGTGTGGGTGTGGGGGAGACAATAAGAGCTGCAC 2048
Db 3284 TGACTCCGGTCTATGAGGACCGCTGTGTGTGGCCAGCAAGCGCTCCACCTGCACA 3225
QY 2049 GAAGTGACTGGACTCTTCCAAAGCCATGCATGCGTACAAATTTGGTGGTCCCATCC 2108
Db 3224 CACCGTGAGCGGGACCTTCAGGAGGCGGAGGCGCTGGGTATGTGGATGTGGGGCTGAT 3165
QY 2109 CGCAGGCGCTCAAGCATCAGCATCCGCAAGCGCGGTACAAAGGCGTGTATCGGGATGA 2168
Db 3164 CCCAGCGGGCA-----CGCAGATCCCGCATCAAGAGGTTGCCGAGGCTGC 3117
QY 2169 CAACTACCTGGCTCTGAAGAACAGCCAAAGCAAGTACCTGCTCAACGGGCAATTCGTGGT 2228
Db 3116 CAACTTCTGGCACTGGGAGTGGAGACCGCGGAGAGTACTTCTCAATGGTGGCTGGAC 3057
QY 2229 GTGGGCGGTGAGCGGACCTGTGTGTAAGGGAGTCTGCTGCGGTACAGCGGCGACGG 2288
Db 3056 CATCCAGTGAAGCGGACTTACCAGGTGGCAGGGACCACTTCACATACGACGACGAGGG 2997
QY 2289 CACAGCGGTGAGAGCTGACAGCTTCCCGCGCCATCCTGGAGCGCTGACCGTGGAGGT 2348
Db 2996 CA---ACTGGGAACCTACGTCCTCCCGGTCCCAAGGAGGCTGTCTGGATCAGCT 2940
QY 2349 CCTCCGTGGGAAGATACACCGCCCGGGTCCGCTACTCTCTCTTA 2396
Db 2939 GCTGTCCAGGAGCAACCTTGGGTGCATAGGATACACATCCA 2892

```

RESULT 14

```

US-09-369-364A-6
; Sequence 6, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 3218
; TYPE: DNA
; ORGANISM: Homo sapiens ADAMTS-7
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13)..(3003)
US-09-369-364A-6

```

```

Query Match 8.4%; Score 240.2; DB 4; Length 3218;
Best Local Similarity 51.4%; Pred. No. 2.2e-43;
Matches 933; Conservative 0; Mismatches 798; Indels 84; Gaps 13;

QY 562 CTGGACCTTTACAGCGCGGCGGGGCTTCGGGGAGAGTCTGAGCGGCGAGTCT 621
Db 646 CTGGAGTCTCGAGCGGAGCGTTCGGAGACGCGGACAGTGGCGGCGGACAGGCTGAGG 705
QY 622 GGGCGGCGCAAGGTTTGTGTCTATCCCGCGGTACGTGGAGACGCTGTGTGGCGGAC 681
Db 706 CGTCTACACCGCGGCTGCTGAGCAAGAGAGTGTGTGAGACCTGTGTAGTGTATGAT 765
QY 682 GAGTCAATGCTCAAGTTCACGG---CGCGGACCTGGACATATATCTGTGAGGCTGTG 738

```

```

Db 766 GCCAAAATGGTGGAGTACCACGACGCGCAGGTTGAGAGCTATGTCTGACATCATG 825
QY 739 GCAAGCGGCGCGGACTCTTACCGCCATCCAGCATCTCTAAACCCCATCAAGATGTTGTG 798
Db 826 AACATGGTGGCTGGGCTGTTCATGACCCAGCATTTGGAAACCCCATCCATCAACATT 885
QY 799 GTCAAGGTGCTCTTCTTAGAGATCGTACCTCCGGGCCAAGGTCAACGGCAATCGGCC 858
Db 886 GTGGCGCTGTCTCTTGGAGATGAGGAGGAGACCTAAAGATCAGGACCATCAGAC 945
QY 859 CTGAGCGTCGCAACCTTCTGTGCTGGCAGAGAAAGTGAACAAAGTGAAGTGAACAGC 918
Db 946 AACACCTTGAAGCTTCTGCAAGTGGCAGAAAGCATCAACATCAAGGGGATGCCCAT 1005
QY 919 CCGGAGTACTGGGACACTGCGCATCTCTTCCACGAGGAGGACCTGTGTGGAGGCCACCA 976
Db 1006 CCCCTGCACCATGACATGCCATCTCTGCTACACGAAAGGACCTGTGTGAGGCCATGAC 1065
QY 977 ----CCTGTGACACCTTGGGCTGCTGTGCTGAGTGGGTACCATGTGTGACCCCAAGAGA 1032
Db 1066 CCGCCCTGTGAGACCTGGGACTGTCCATGTGGCGGGCATGTGCCACGCGGACCGCAGC 1125
QY 1033 TGCTCTGTCATTTGAGGACGATGGGCTTCCATAGCCTTTCACCACTGCCACGAGTGGGC 1092
Db 1126 TGCAGCATCAACGAGGACGACGCGGCTGCGCTGGCTTCACTGTAGCCACGAGCTCGG 1185
QY 1093 CACGTGTTCAACATGCCCATGACAAATGTGAAGTCTGTGAGGAGGTGTTTGGAGCTC 1152
Db 1186 CACAGTTTGGCATTCAGCATGACGAAAGCGGCAATGACTGTGAGCCGCTTGTGGAAACGA 1245
QY 1153 CGAGCCAAACCATGATGTCCCGCCCTTCCATCCAGATGACCGGTGCCAACCCCTGGTCA 1212
Db 1246 CTTTTCATC-----ATGTCTCCACGCTCTGTGACGAGCGGCTCCCTCACTGGTCC 1299
QY 1213 GCTGAGTGTGTCATCATCAGCTTCTTGGACAGCGGCGGAGTGTGCTCTCTCTCTCT 1272
Db 1300 CGCTGACGCGGCGAGTATATACCAAGTTCTTGTACCGGTGGGTGGGCTGTGCTGGAC 1359
QY 1273 GACCAACACGACCAAGCC---ATCTCCCTGCGGAGGATCTCCGGGCGCAGCTACACC 1329
Db 1360 GACCTCTTGCACGAGCATTTATCGACTTCCCTCGGTGGCCACCTGGGCTCTCTATGAT 1419
QY 1330 CTGAGCCAGAGTGGCGAGTGGCTTTTGGCGTGGCTCCAAAGCCCTGTCTTACATGAG 1389
Db 1420 GTAAGCCACAGTGGCGCTCCAGTACGCGGCTACTCTGCTCTTCTGCGAGGACATGAT 1479
QY 1390 TA---CTGACCAACGCTGTGTGACCGGAGGCGGACGAGGACAGATGTTGTGCGAGCC 1446
Db 1480 AATGTCTGCCACACACTCTGTGTCTGTGGGGACCA-----CCTGTACATCC 1527
QY 1447 CGCCACTTCCCTGGGCGGATGGCACAGCTGTGGGAGGCAAGCTCTGCTCAAGGG 1506
Db 1528 AAGCTGGATGAGCTGTGGACGCGGACCGGCTGTGGGAGAAATAGTGTCTCTAGTGG 1587
QY 1507 GCTGCGTGGAGAGACAACTTCAACAGCACAGGCTGGATGTTCTTGGGCCCAATGG 1566
Db 1588 GAGTGGTACCGCTGGGCTTCCGGCCGAGGCG---GTGGATGTTGGCTGTGTGGCTGG 1644
QY 1567 GATCCCTATGGCCCTGTCTGGGCAATGTGTGGGGCGGTGACGCTGGCCAGAGGAG 1626
Db 1645 AGCGCTGTGTCATCTGTCTACGAGCTGTGGCATGGGCTGACAGGCGGAGCGGAG 1704
QY 1627 TGCACAAACCCACCCCTGCCAACGGGGCAAGTACTCGAGGAGGTGAGGTGAGGATAC 1686
Db 1705 TGCACGAGCTTACGCCCAATACAAAGGAGATACTGTGTGGTGAGCGCAAGCGCTTC 1764
QY 1687 CGATCTCTCAACTGGAGCCCTGCCCCAGCTCAGCTCCGGAAAGAGCTTCCGGGAGGAG 1746
Db 1765 CGCCTCTCAACTGACGCGCTGCC-----CTGCTGSCCGCCCTCTCTCCGCCACGTC 1818
QY 1747 CAGTGTGAGGCTTTCACAGGCTTACACCCAGCAGCAGCAGGCTCACTCTCGCGTGGCA 1806
Db 1819 CAGTGCAGCCACTTGTGAGCTATGCTCTACAAGGGCGGAGCTGCACA-----CA 1866

```

QY 1807 TGGGTGCCAAGTACTCCGCGTGTCTCCCGGGAACAAGTGAAGCTCATCTGCCGAGCC 1866
 Db 1867 TGGGTGCCGCTGTCTCAATGACGTGAACCC-----TGCAGCTGCACCTGCCGCGCC 1917
 QY 1867 AATGGCACTGGCTACTTCTATGTCTGGGACCAAGTGTGGACGCGCTGTCTCT 1926
 Db 1918 GCGAATGAGTACTTTGCAAGAAAGTCTGGGACGCTGTCTGATGGCACCCCTCTCTAC 1977
 QY 1927 CCGTACTCCACC-----TCCGCTGTGTCCAAAGGAAGTGCATCAAGGCTGCTGTGAT 1980
 Db 1978 CAGTCCGAGCGCAGCCGGGACCTCTGCATCAACGGCATCTGTGAAGACGTGGGCTGTGAC 2037
 QY 1981 GGGAACTGGGCTCCRAAGAAGATTGCAACAGTGTGGGCTGTGTGGGGAGACAATAAG 2040
 Db 2038 TTCGAGATTGACTCCGCTGTATGGAGGACCGCTGTGTGTGGCCACGCAACGGCTCC 2097
 QY 2041 AGCTGCAAGAGGTGACTGACTCTTCAACCAAGCCCATGATGGCTGACATTTTCGTGGTG 2100
 Db 2098 ACCTGCCACACCGTGTAGCGGGACCTTCGAGGAGGCGGAGGCTTGGGGTATGTGGATGTG 2157
 QY 2101 GCCATCCCGCAGCGGCTCAAGCATCGACATCCGCGAGCGGTTACAAAGGCTGATC 2160
 Db 2158 GGGCTGATCCAGCGGCGCA-----CGCAGATCCGCATCCAGAGGTGGC 2205
 QY 2161 GGGGACACACTACTCGCTCTGAAGAACAGCAAGGCAAGTACCTGCTCAACGGGCAAT 2220
 Db 2206 GAGGCTGCCAACTTCCTTGGCACCTGGGAGGAGGAGCCCGGAGAAGTACTTCTCAATGGT 2265
 QY 2221 TTCGTGTGTGGGGTGGAGCGGGACCTGGTGGTGAAGGGCAAGTCTGTGCGGGTACAGC 2280
 Db 2266 GGCTGGACCATCCAGTGGAAAGCGGAGTACACAGGTGGCAGGGACCACTTTCATATACGCA 2325
 QY 2281 GGCACGGGACAGCGGTGGAGAGCTTCAGGCTTCGCGGCCCATCTGGAGCGGCTGACC 2340
 Db 2326 CGCAGGGGCA---ACTGGGAGAACTCACGTCCCGGGTCCCAAGGAGCGCTGTCTGG 2382
 QY 2341 GTGGAGTCTCTCC 2355
 Db 2383 ATCCAGGTGCTGCC 2397

RESULT 15
 US-09-369-364A-3
 ; Sequence 3, Application US/09369364A
 ; Patent No. 6391610
 ; GENERAL INFORMATION:
 ; APPLICANT: Apte, Suneel
 ; APPLICANT: Hurskainen, Tiina L.
 ; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
 ; FILE REFERENCE: 26473/4007/10-30-00
 ; CURRENT APPLICATION NUMBER: US/09/369,364A
 ; CURRENT FILING DATE: 1999-08-06
 ; NUMBER OF SEQ ID NOS: 31
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 3
 ; LENGTH: 1520
 ; TYPE: DNA
 ; ORGANISM: homo sapiens ADAMTS-5
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (296)
 ; OTHER INFORMATION: n=T
 ; NAME/KEY: misc_feature
 ; LOCATION: (1443)
 ; OTHER INFORMATION: n=A
 ; NAME/KEY: misc_feature
 ; LOCATION: (1457)
 ; OTHER INFORMATION: n=G
 ; NAME/KEY: misc_feature
 ; LOCATION: (1458)
 ; OTHER INFORMATION: n =T

US-09-369-364A-3

Query Match 7.6%; Score 215.8; DB 4; Length 1520;
 Best Local Similarity 56.0%; Pred. No. 3.7e-38;
 Matches 456; Conservative 0; Mismatches 347; Indels 12; Gaps 2;
 QY 1603 GCGGTGACGTGGCCAGGAGGAGTGCACCAACCCACCCCTGCCAAGCGGGGGAAGTAC 1662
 Db 360 GGAGTGCAGTTTGGCTATCGTCACTGTATACCCCTGCTCCAGAAACACAGCGCTAC 419
 QY 1663 TCGGAGGAGTGAAGGTGAATACCATCTGCAACCTGGAGCGCTGCCCGCAGCTCAGCC 1722
 Db 420 TGCACAGGAAGAGGCGCATCTACCACTCCCTGCACTCATGTCCTCCCTGCC-----CACCC 473
 QY 1723 TCCGGAAGAGAGTTCCTGGGAGGAGCAGTGTGAGGCTTTTCAACGGCTCAACACAGCAC 1782
 Db 474 AATGTAATCATTTTCGTGCATCAACAGTGTGAGGCCAAATATGCTATCACTCTATGCA 533
 QY 1783 AACCGGCTCACTCTCGCGTGGCATGGTGGTCCCAAGTACTCCGGGCTGTCTCCCGGGGAC 1842
 Db 534 AAAGAGTCAAACTTTTGTGAATGGTTCCTCAAAATATGCAAGTGTCTCTGCCAGCGAT 593
 QY 1843 AAGTCAAGCTCATCTCGCGAGGCCAATGCACTGGCTACTTCTATGTCTCGCACCCCAAG 1902
 Db 594 GTGTGCAAGCTGACCTGCAGAGCCCAAGGCACTGGCTACTATGTGGTATTTTC7CAAAG 653
 QY 1903 GTGTGGGAGCGCACGCTGTGCTCTCTGACTCCACCTCCGCTGTGTGTCTTCC7AGGCAAGTGC 1962
 Db 654 GTGACCGATGGCACTGAATGTAGGCGGTACAGTAATTCCTGCTCGGGGGAAGTGT 713
 QY 1963 ATCAAGGTGGTGTGTATGGGAACCTGGGTCCCAAGAGAGATTCGCAAGTGTGGGGTG 2022
 Db 714 GTGAGAAGTGGTGTGAGCGCATCATTTGGCTCAAAAGTGTGCAAGTGTGCAAGTGGGAGTA 773
 QY 2023 TGTGGGAGACAATAAGAGCTGCAAGAGTGTGACTGCTTTCACCAAGCCCATGAT 2082
 Db 774 TGTGGAGAGACATCCAGCTGTACAAGATTTTGGACCTTTATAGAAAGTGAAG 833
 QY 2083 GGCTACAATTTCTGTGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCCGCGAGCGC 2142
 Db 834 GGTACACTGACGTGTGTGAGGATTCCTGAAGGGGCAACCCACATAAAAGTTCGACAGTTC 893
 QY 2143 GGTACAAAGGCTGATCGGGATGACAACCTACCTGGCTCTGAAGAACAGCAAGGCAAG 2202
 Db 894 AAAGCCAAAGACCAAGACTAGATTCACTGCTATTTAGCCCTGAAAAGAAAACGGGTGAG 953
 QY 2203 TACCTGCTCAACGGCATTTCTGTGTGTGGCGGTGGAGCGGACCTGGTGGTGAAGGCG 2262
 Db 954 TACCTTATCAATGGAAAGTACATGATCTCCACTTCAGAGACTATCATTTGACATCAATGGA 1013
 QY 2263 AGTCTGCTGCGGTACAGCGGACAGCGGACAGCGGTGGAGAGCCTGCA-----GGCTTCC 2316
 Db 1014 ACAGTCAATGAACATATAGCGGTGGAGCCACAGGATGACTTCTCGATGCGATGGGCTAC 1073
 QY 2317 CGGCCCCCTCGGAGCGCTGACCGTGGAGGTCTCTCCGTGGGGAAGATGAC7CCGCCC 2376
 Db 1074 TCTGCCCAAGGAATTTCTAATAGTGCAGATTCTTGAACAGACCCCACTAAACCATTA 1133
 QY 2377 CGGGTCCGCTACTCTCTCTATCTGCCCCAAAGGCC 2411
 Db 1134 GATGCGGTATAGCTTTTGTTCCTCCAAAGATC 1168

Search completed: May 1, 2003, 22:52:44
 Job time : 162 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2003, 19:41:41 ; Search time 339 Seconds
(without alignments)
9944.440 Million cell updates/sec

Title: US-10-009-332-2
Perfect score: 2853
Sequence: 1 atgcttttctgggcatcct.....gcgtcctgagccgtgctga 2853

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published_Applications_NA:*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
 - 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
 - 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
 - 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2846.6	99.8	2853	10 US-09-965-631-3	Sequence 3, Appli
2	2846.6	99.8	3446	10 US-09-965-631-7	Sequence 7, Appli
3	2296.4	80.5	2469	9 US-10-163-316-3	Sequence 3, Appli
4	2296.4	80.5	2940	9 US-10-163-316-1	Sequence 1, Appli
5	1088	38.1	1104	10 US-09-965-631-5	Sequence 5, Appli
6	955.8	33.5	966	10 US-09-965-631-1	Sequence 1, Appli
7	636	22.3	2184	9 US-10-097-597-13	Sequence 13, Appli
8	636	22.3	2184	9 US-10-097-580-13	Sequence 13, Appli
9	636	22.3	2184	9 US-10-097-580-2	Sequence 2, Appli
10	628.8	22.0	2184	9 US-10-097-580-2	Sequence 2, Appli
11	628.8	22.0	2184	10 US-09-445-023A-2	Sequence 2, Appli
12	628.2	22.0	4676	12 US-10-105-929-1	Sequence 1, Appli
13	628.2	22.0	3638	10 US-09-918-171A-8	Sequence 8, Appli
14	592	20.8	4407	9 US-10-174-590-351	Sequence 351, App
15	583.2	20.4	4407	9 US-10-176-758-351	Sequence 351, App
16	583.2	20.4	4407	9 US-10-175-737-351	Sequence 351, App
17	583.2	20.4	4407	9 US-10-173-706-351	Sequence 351, App
18	583.2	20.4	4407	9 US-10-173-697-351	Sequence 351, App
19	583.2	20.4	4407	9 US-10-173-697-351	Sequence 351, App

ALIGNMENTS

RESULT 1
US-09-965-631-3
; Sequence 3, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Enco
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-965-631-3

Query Match	99.8%	Score 2846.6;	DB 10;	Length 2853;
Best Local Similarity	99.9%	Pred. No. 0;		
Matches 2849;	Conservative	0; Mismatches	4; Indels	0; Gaps
0;				
Qy	1	ATGCTTTTCTGGGCATCCTAACCCCTGGCTTCGCGGGGGAACCGCTGGAGGCTTTGAG	60	
Db	1	ATGCTTCTGCTGGGCATCCTAACCCCTGGCTTCGCGGGGGAACCGCTGGAGGCTTTGAG	60	
Qy	61	CCAGAGGGGAGGTAGTCTGCTTCCATCCGACTGGACCCGGGACATTAACGCTCGCTAC	120	
Db	61	CCAGAGGGGAGGTAGTCTGCTTCCATCCGACTGGACCCGGGACATTAACGCTCGCTAC	120	
Qy	121	TACTGGGGGGTCCCGAGGACTCCGGGGATCAGGACTCATTTTTCAGATCAGACATTT	180	
Db	121	TACTGGGGGGTCCCGAGGACTCCGGGGATCAGGACTCATTTTTCAGATCAGACATTT	180	
Qy	181	CAGGAGGACTTTTACCTACACCTTGACCGCGGATGCTCAGTCTTGGCTCCGCTTCTCC	240	
Db	181	CAGGAGGACTTTTACCTACACCTTGACCGCGGATGCTCAGTCTTGGCTCCGCTTCTCC	240	
Qy	241	ACTGAGCATCTGGGCGTCCCGCTCCACGGGGCTCACCGGGGGCTCTTTCAGTCTTGGC	300	

[illegible]


```
QY 2461 TTGCACAAACAGCTCTCTCAGCCCTTCCAAACAGGTGAGCAGCCGCGACAGGCCCTT 2520
|
|
|
Db 2461 TTGCACAAACAGCTCTCTCAGCCCTTCCAAACAGGTGAGCAGCCGCGACAGGCCCTT 2520
|
|
|
QY 2521 GCACGCTGGGTGCTGGCAGCTGGGGCGCTGTCCGCGAGCTGCGGCGAGTGGCCTGCGAG 2580
|
|
|
Db 2521 GCACGCTGGGTGCTGGCAGCTGGGGCGCTGTCCGCGAGCTGCGGCGAGTGGCCTGCGAG 2580
|
|
|
QY 2581 AAGCGGCGGTGACGTCCGCGGCTCCGCGGCGCAGCGCAGGTCCCTGCTGTGATGCA 2640
|
|
|
Db 2581 AAGCGGCGGTGACGTCCGCGGCTCCGCGGCGCAGCGCAGGTCCCTGCTGTGATGCA 2640
|
|
|
QY 2641 GCCCATCGGCCCTGGAGACACAAAGCTGCGGGGAGCCCTGCCCACTGGGAGCTCAGC 2700
|
|
|
Db 2641 GCCCATCGGCCCTGGAGACACAAAGCTGCGGGGAGCCCTGCCCACTGGGAGCTCAGC 2700
|
|
|
QY 2701 GCCTGTGTCACCTGCTCCAGAGCTGCGGGCGGGGATTTTCAGAGGGCTTCACTCAAGTGT 2760
|
|
|
Db 2701 GCCTGTGTCACCTGCTCCAGAGCTGCGGGCGGGGATTTTCAGAGGGCTTCACTCAAGTGT 2760
|
|
|
QY 2761 GTGGGCCACGAGCGCGCTGTGCGCGGCGGACAGTCACTTGCACCGCAGCCCGAG 2820
|
|
|
Db 2761 GTGGGCCACGAGCGCGCTGTGCGCGGCGGACAGTCACTTGCACCGCAGCCCGAG 2820
|
|
|
QY 2821 GAGCTGGACTTCTGCGCTCTGAGGCGGCTGCTGA 2853
|
|
|
Db 2821 GAGCTGGACTTCTGCGCTCTGAGGCGGCTGCTGA 2853
|
|
|
```

RESULT 2

```
US-09-965-631-7
; Sequence 7, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Hilbun, Erin
; APPLICANT: Fridde, Carl Johan
; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Encodin
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 3446
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-965-631-7
```

```
Query Match 99.8%; Score 2846.6; DB 10; Length 3446;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2849; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 ATGCTTTCTGCTGGGATCTTAACCTGGCTTTTCGGCGGCGAAGCGTGGAGGCTTTGAG 60
|
|
|
Db 397 ATGCTTTCTGCTGGGATCTTAACCTGGCTTTTCGGCGGCGAAGCGTGGAGGCTTTGAG 60
|
|
|
QY 61 CCAGAGCGGAGGTAGTCTTCCATCCGACTGGACCGGACATTAAGCGCGCGCTTAC 120
|
|
|
Db 457 CCAGAGCGGAGGTAGTCTTCCATCCGACTGGACCGGACATTAAGCGCGCGCTTAC 120
|
|
|
QY 121 TACTGGCGGGGTCGCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCACACATTT 180
|
|
|
Db 517 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCACACATTT 180
|
|
|
QY 181 CAGGAGGACTTTTACTACACCTGACGCGGATGCTAGTCTTGGCTCCGCGCTTCTCC 240
|
|
|
Db 577 CAGGAGGACTTTTACTACACCTGACGCGGATGCTAGTCTTGGCTCCGCGCTTCTCC 240
|
|
|
QY 241 ACTGAGCATCTGGGCTCCCGCTCCAGGGCTTCACGGGGCTCTTCAGACCTGGAGCG 300
|
|
|
```

```
Db 637 ACTGAGCATCTGGGGCTCCCGCTCCAGGGGCTCACCAGGGGCTCTTTCAGTCTCTCGAGCG 696
|
|
|
QY 301 TGCCTTCTATTCTGGGGAGCTGAACCGCGAGCGGACTCGTTTCGGTGTGTGAGCTGTGCG 360
|
|
|
Db 697 TGCCTTCTATTCTGGGGAGCTGAACCGCGAGCGGACTCGTTTCGGTGTGTGAGCTGTGCG 756
|
|
|
QY 361 GGGGGGCTCCGCGGAGCTTTGGCTACCGAGCGCGGAGTATGTCATTAGCCCGCTGCC 420
|
|
|
Db 757 GGGGGGCTCCGCGGAGCTTTGGCTACCGAGCGCGGAGTATGTCATTAGCCCGCTGCC 816
|
|
|
QY 421 AATGCTAGCGCGCGCGCGCAGCGCAACAGCCAGGCGCAGACCTTCTCCAGCGCGG 480
|
|
|
Db 817 AATGCTAGCGCGCGCGCGCAGCGCAACAGCCAGGCGCAGACCTTCTCCAGCGCGG 876
|
|
|
QY 481 GGTGTTCCGGGGGCGCTTCCGAGAGACCCACCTCTCGCTCGGGGTGGCTCGGGGTGG 540
|
|
|
Db 877 GGTGTTCCGGGGGCGCTTCCGAGAGACCCACCTCTCGCTCGGGGTGGCTCGGGGTGG 936
|
|
|
QY 541 AACCCCGCATCTTACGGGCGCTTGACCCCTTACAAGCGCGCGGCGGGGCTTCCGGGAG 600
|
|
|
Db 937 AACCCCGCATCTTACGGGCGCTTGACCCCTTACAAGCGCGCGGCGGGGCTTCCGGGAG 996
|
|
|
QY 601 AGTCGTAGCGCGCGGAGCTTGGGCGCGCCAAAGCTTTTCGTGTCTATCCCGCGCTAGCTG 660
|
|
|
Db 997 AGTCGTAGCGCGCGGAGCTTGGGCGCGCCAAAGCTTTTCGTGTCTATCCCGCGCTAGCTG 1056
|
|
|
QY 661 GAGAGCTGTGTGCGGAGCAGTCAATGTCAAGTTCCAGCGCGCGGCGGACCTGGACAT 720
|
|
|
Db 1057 GAGAGCTGTGTGCGGAGCAGTCAATGTCAAGTTCCAGCGCGCGGAGCCTGGACAT 1116
|
|
|
QY 721 TATCTGCTGAGCGCTGCTGGCAACGCGCGGCGGACTCTACCGCCATCCAGCATCTCTCAAC 780
|
|
|
Db 1117 TATCTGCTGAGCGCTGCTGGCAACGCGCGGCGGACTCTACCGCCATCCAGCATCTCTCAAC 1176
|
|
|
QY 781 CCCATCAACATCGTTGGTCAAGGTGCTGCTTCTAGAGATCGTACTCGGGGCCAAG 840
|
|
|
Db 1177 CCCATCAACATCGTTGGTCAAGGTGCTGCTTCTAGAGATCGTACTCGGGGCCAAG 1236
|
|
|
QY 841 GTCACCGCATGCGGCGCTGACCTGCGCAACTTCTGTGCTGGCAGAGAACTCAAC 900
|
|
|
Db 1237 GTCACCGCATGCGGCGCTGACCTGCGCAACTTCTGTGCTGGCAGAGAACTCAAC 1296
|
|
|
QY 901 AAAGTGTGAGTGAACAGCAGTGTGGGAGTGTGCTTCTTGTAGATCGTACTCGGGGCCAAG 960
|
|
|
Db 1297 AAAGTGTGAGTGAACAGCAGTGTGGGAGTGTGCTTCTTGTAGATCGTACTCGGGGCCAAG 1356
|
|
|
QY 961 CTGTGTGGAGCCACCTGTGACACCGTGGGATGCTGATGGGTACATGTGTGAC 1020
|
|
|
Db 1357 CTGTGTGGAGCCACCTGTGACACCGTGGGATGCTGATGGGTACATGTGTGAC 1416
|
|
|
QY 1021 CCCAAGAGAAGCTGCTGTCTATTGAGAGCATGGGCTTCCATCAGCTTCACTGTC 1080
|
|
|
Db 1417 CCCAAGAGAAGCTGCTGTCTATTGAGAGCATGGGCTTCCATCAGCTTCACTGTC 1476
|
|
|
QY 1081 CACGAGCTGGGCCAGCTGTTCAACATGCCCATGCCCATGACAATGTGAAAGTCTGTGAGAGGTG 1140
|
|
|
Db 1477 CACGAGCTGGGCCAGCTGTTCAACATGCCCATGCCCATGACAATGTGAAAGTCTGTGAGAGGTG 1536
|
|
|
QY 1141 TTTGGGAGCTCCGAGCAACCATCATGTCCCGGACCTTCATCCAGATCGACCGTGC 1200
|
|
|
Db 1537 TTTGGGAGCTCCGAGCAACCATCATGTCCCGGACCTTCATCCAGATCGACCGTGC 1596
|
|
|
QY 1201 AACCCCTGGTCCAGCTGCTGAGTGTGCCATCATCACCGACTTCTCTGGACAGCGGCGAGGT 1260
|
|
|
Db 1597 AACCCCTGGTCCAGCTGCTGAGTGTGCCATCATCACCGACTTCTCTGGACAGCGGCGAGGT 1556
|
|
|
QY 1261 GACTGCTCTCTGGACCAACAGCAAGCCCATCTCTCCCTGCCCGAGGATCTGCGGGGCGCC 1320
|
|
|
Db 1657 GACTGCTCTCTGGACCAACAGCAAGCCCATCTCTCCCTGCCCGAGGATCTGCGGGGCGCC 1716
|
|
|
QY 1321 AGCTACACCTTGAGCCAGCAGTGGCTGCTTTTGGCGTGGGCTCCCAAGCCTCTCT 1380
|
|
|
Db 1717 AGCTACACCTTGAGCCAGCAGTGGCTGCTTTTGGCGTGGGCTCCCAAGCCTCTCTCT 1776
|
|
|
```

QY 1381 TACATGACGAGTACGACCAAGCTGTGGTGACCGGAAGCCCAAGGACAGATGGTGTGC 1440
 DB 1777 TACATGACGAGTACGACCAAGCTGTGGTGACCGGAAGCCCAAGGACAGATGGTGTGC 1836
 QY 1441 CAGACCGCGCATTCCCTGGGCGGATGGACACAGCTGTGGGAGGAGCAAGCTGTGCCTC 1500
 DB 1837 CAGACCGCGCATTCCCTGGGCGGATGGACACAGCTGTGGGAGGAGCAAGCTGTGCCTC 1896
 QY 1501 AAGGGGCTTGGTGGAGAGACACACCTCAACACACAGAGTGTGATGGTTCCTGGGCC 1560
 DB 1897 AAGGGGCTTGGTGGAGAGACACACCTCAACACACAGAGTGTGATGGTTCCTGGGCC 1956
 QY 1561 AATATGGATCCCTATATGGCCCTCTCTCGGCACATGTGGTGGGCGCTGCAGCTGGCCAGG 1620
 DB 1957 AATATGGATCCCTATATGGCCCTCTCTCGGCACATGTGGTGGGCGCTGCAGCTGGCCAGG 2016
 QY 1621 AGGAGTGCACCAACCCACCTCTGCCAAGGGGGCAAGTACTGCGAGGAGTGAAGGTG 1680
 DB 2017 AGGAGTGCACCAACCCACCTCTGCCAAGGGGGCAAGTACTGCGAGGAGTGAAGGTG 2076
 QY 1681 AATATCCGATCCCTGACACCTGGAGCCCTGCCACAGCTCAGCTCCGGAAGAGCTTCCGG 1740
 DB 2077 AATATCCGATCCCTGACACCTGGAGCCCTGCCACAGCTCAGCTCCGGAAGAGCTTCCGG 2136
 QY 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACACCAACAGCAGCAACCGGCTCACTCTCGCC 1800
 DB 2137 GAGGAGCAGTGTGAGGCTTTCAACGGCTACACCAACAGCAGCAACCGGCTCACTCTCGCC 2196
 QY 1801 GTGCATGGTGGCCCAAGTACTCGGGGTGTCTCCCGGGACAAAGTGCAGCTCATCTGC 1860
 DB 2197 GTGCATGGTGGCCCAAGTACTCGGGGTGTCTCCCGGGACAAAGTGCAGCTCATCTGC 2256
 QY 1861 CGACCAATGGCACTGCTACTTCTATGTCTGGCACCCCAAGTGTGGACGCAAGCTG 1920
 DB 2257 CGACCAATGGCACTGCTACTTCTATGTCTGGCACCCCAAGTGTGGACGCAAGCTG 2316
 QY 1921 TGCTCTCTGACTGCTGCTGCTGTGTGTCCAAAGGCAAGTGCATCAAGGCTGCTGTGAT 1980
 DB 2317 TGCTCTCTGACTGCTGCTGCTGTGTGTCCAAAGGCAAGTGCATCAAGGCTGCTGTGAT 2376
 QY 1981 GGGAACTGGCTTCCAAAGAGATTCGCAAGTGTGGGTGTGGGGGAGACAATAAG 2040
 DB 2377 GGGAACTGGCTTCCAAAGAGATTCGCAAGTGTGGGTGTGGGGGAGACAATAAG 2436
 QY 2041 AGCTGCAAGAGTGTGACTGCTGCTTCCCAAGCCATGCTGCTACATTTCTGCTGTG 2100
 DB 2437 AGCTGCAAGAGTGTGACTGCTGCTTCCCAAGCCATGCTGCTACATTTCTGCTGTG 2496
 QY 2101 GCCATCCCGGAGGCGCTCAAGCATCGACATCGCCAGCGCGGTTACAAGGGCTGATC 2160
 DB 2497 GCCATCCCGGAGGCGCTCAAGCATCGACATCGCCAGCGCGGTTACAAGGGCTGATC 2556
 QY 2161 GGGATGACAACTACCTGGCTCTGAAGACAGCAAGGCAAGTACCTGCTCAACGGGCAT 2220
 DB 2557 GGGATGACAACTACCTGGCTCTGAAGACAGCAAGGCAAGTACCTGCTCAACGGGCAT 2616
 QY 2221 TTCGTGTGTGGCGGTGGAGCGGACCTGTGTGTGAAGGCAAGTGTGTGCGGTACAGC 2280
 DB 2617 TTCGTGTGTGGCGGTGGAGCGGACCTGTGTGTGAAGGCAAGTGTGTGCGGTACAGC 2676
 QY 2281 GGCACGGGCACAGGGTGGAGAGCTGACAGCTTCCCGGCCCATCTCGGAGCGCTGACC 2340
 DB 2677 GGCACGGGCACAGGGTGGAGAGCTGACAGCTTCCCGGCCCATCTCGGAGCGCTGACC 2736
 QY 2341 GTGAGGTCTCTCTCGGTGGGAAGTACACCGCCCGGGTCCGCTACTCTTCTATCTG 2400
 DB 2737 GTGAGGTCTCTCTCGGTGGGAAGTACACCGCCCGGGTCCGCTACTCTTCTATCTG 2796
 QY 2401 CCCAAAGAGCTCGGGAGGACAAAGTCTCTCATCCCAAGGACCCCGGGGACCCCTGTGC 2460
 DB 2797 CCCAAAGAGCTCGGGAGGACAAAGTCTCTCATCCCAAGGACCCCGGGGACCCCTGTGC 2856

QY 2461 TTGCACAAACAGCGTCTCAGCCTCTCCAAACAGGTGGAGCAGCGGAGCAGAGCCGCCCT 2520
 DB 2857 TTGCACAAACAGCGTCTCAGCCTCTCCAAACAGGTGGAGCAGCGGAGCAGAGCCGCCCT 2916
 QY 2521 GCACGCTGGTGGCTGGCAGCTGGGGCGGCTCCGCGAGCTGGCAGTGGCGCTGCAG 2580
 DB 2917 GCACGCTGGTGGCTGGCAGCTGGGGCGGCTCCGCGAGCTGGCAGTGGCGCTGCAG 2976
 QY 2581 AAGCGGCGGTGGACTGCGGGGGTCCCGGGGAGCGACGCGTCCCTGTGTGATGCA 2640
 DB 2977 AAGCGGCGGTGGACTGCTGGGGCTCCCGGGGAGCGACGCGTCCCTGTGTGATGCA 3036
 QY 2641 GCCATCGGCGGTGGAGACACAAAGCTGCGGGGAGCGCTGCCCGACCTGGAGCTCAGC 2700
 DB 3037 GCCATCGGCGGTGGAGACACAAAGCTGCGGGGAGCGCTGCCCGACCTGGAGCTCAGC 3096
 QY 2701 GCCTGGTCACTGCTTCCAAAGCTGCGGGGAGCGGATTTTCAGAGCGCTCACTCAAGTGT 2760
 DB 3097 GCCTGGTCACTGCTTCCAAAGCTGCGGGGAGCGGATTTTCAGAGCGCTCACTCAAGTGT 3156
 QY 2761 GTGGGCCACGAGCGCGCTGCTGGCCCGGAGCAGTGCACCTTGCACCGCAAGCCCGCAG 2820
 DB 3157 GTGGGCCACGAGCGCGCTGCTGGCCCGGAGCAGTGCACCTTGCACCGCAAGCCCGCAG 3216
 QY 2821 GAGCTGGACTTCTGCGTCTGAGGCGCTGCTGA 2853
 DB 3217 GAGCTGGACTTCTGCGTCTGAGGCGCTGCTGA 3249

RESULT 3
 US-10-163-316-3
 ; Sequence 3, Application US/10163316
 ; Publication No. US20020197703A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kapeller-Libermann, Rosana
 ; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
 ; TITLE OF INVENTION: Therefor
 ; FILE REFERENCE: MP101-025P1RNM
 ; CURRENT APPLICATION NUMBER: US/10/163,316
 ; CURRENT FILING DATE: 2002-06-05
 ; PRIOR APPLICATION NUMBER: 60/297,863
 ; PRIOR FILING DATE: 2001-06-13
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 2469
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(2469)
 ; US-10-163-316-3

Query Match 80.5%; Score 2296.4; DB 9; Length 2469;
 Best Local Similarity 96.9%; Pred. No. 0;
 Matches 2379; Conservative 0; Mismatches 6; Indels 69; Gaps 1;

QY 1 ATGCTTTTCTGGGATCTCACTTACCTGCTTCGCGGGCGGAGCGCTGGAGGCTTTCAG 60
 DB 1 ATGCTTCTGCTGGCATCTCACTTACCTGCTTCGCGGGCGGAGCGCTGGAGGCTTTCAG 60
 QY 61 CCAGAGCGGAGGTAGTCTTCCATCCGACTGGACCGGACATTAAGCGCGCGCTAC 120
 DB 61 CCAGAGCGGAGGTAGTCTTCCATCCGACTGGACCGGAGCATTAAGCGCGCGCTAC 120
 QY 121 TACTGCGGGGTCCCGAGGACTCCGGGGATCAGGAGCTCAATTTTTCAGATCAGACATTT 180
 DB 121 TACTGCGGGGTCCCGAGGACTCCGGGGATCAGGAGCTCAATTTTTCAGATCAGACATTT 180
 QY 181 CAGGAGGACTTTTACCTACACCTGACCGGGATGCTCAGTCTTGGCTCCGCGCTTCC 240
 DB 181 CAGGAGGACTTTTACCTACACCTGACCGGGATGCTCAGTCTTGGCTCCGCGCTTCC 240

QY 241 ACTGAGCATGCGCGTCCCTCCAGGGCTCACGGGGCTTTCAGACCTGCGACGC 300
DB 241 ACTGAGCATGCGCGTCCCTCCAGGGCTCACGGGGCTTTCAGACCTGCGACGC 300
QY 301 TGCCTCTATCTCTGGGACGTTGACCGCGAGCGGACTGCTTCCGCTGCTGAGCGCTGTC 360
DB 301 TGCCTCTATCTCTGGGACGTTGAAACGCGAGCGGACTGCTTCCGCTGCTGAGCGCTGTC 360
QY 361 GGGGGCTCCGCGAGGCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCGCTGCC 420
DB 361 GGGGGCTCCGCGAGGCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCGCTGCC 420
QY 421 AATGCTAGCGCGCGGCGCGAGCGCAACAGCCAGGGCGCACACTTCTCCAGCGCG 480
DB 421 AATGCTAGCGCGCGGCGCGAGCGCAACAGCCAGGGCGCACACTTCTCCAGCGCG 480
QY 481 GGTGTTCCGGCGGCGCTTCCGAGACCCACCTCTCGCTGCGGGGTGGCTCGGGCTGG 540
DB 481 GGTGTTCCGGCGGCGCTTCCGAGACCCACCTCTCGCTGCGGGGTGGCTCGGGCTGG 540
QY 541 AACCCGCCATCTTACGGGCCCTGGACCCCTTACAAGCGCGCGGCGGCTTCCGGGAG 600
DB 541 AACCCGCCATCTTACGGGCCCTGGACCCCTTACAAGCGCGCGGCGGCTTCCGGGAG 600
QY 601 AGTCGTAGCGCGCGAGGCTGCGGCGCGCAAGCGTTTCTGCTATCCGCGGTACGTG 660
DB 601 AGTCGTAGCGCGCGAGGCTGCGGCGCGCAAGCGTTTCTGCTATCCGCGGTACGTG 660
QY 661 GAGACGCTGTGTCGCGGACGAGTCAATGGTCAAGTCCAGCGCGGACCTGGACAT 720
DB 661 GAGACGCTGTGTCGCGGACGAGTCAATGGTCAAGTCCAGCGCGGACCTGGACAT 720
QY 721 TATCTGCTGACGCTGTGCGCAAGCGGCGGCGACTTACCGGCATCCAGCATCTCTCAAC 780
DB 721 TATCTGCTGACGCTGTGCGCAAGCGGCGGCGACTTACCGGCATCCAGCATCTCTCAAC 780
QY 781 CCATCAACATCTGTTGCTCAAGTGTCTTCTTAGAGATCGTGACTCCGGGCGGCAAG 840
DB 781 CCATCAACATCTGTTGCTCAAGTGTCTTCTTAGAGATCGTGACTCCGGGCGGCAAG 840
QY 841 GTCACGGCAATCGGCGCTGACGCTGCGCAACTTCTGCTGGCGAGAGAGCTGAAC 900
DB 841 GTCACGGCAATCGGCGCTGACGCTGCGCAACTTCTGCTGGCGAGAGAGCTGAAC 900
QY 901 AAAGTGAAGCAAGCAACCGGAGTACTGGGACACTGCCATCCTCTTCCAGGCGAGGAC 960
DB 901 AAAGTGAAGCAAGCAACCGGAGTACTGGGACACTGCCATCCTCTTCCAGGCGAGGAC 960
QY 961 CTGTGTGAGGCCACCACTGTGACACCTGGGCATGGCTGATGGGTACCATGTGTGAC 1020
DB 961 CTGTGTGAGGCCACCACTGTGACACCTGGGCATGGCTGATGGGTACCATGTGTGAC 1020
QY 1021 CCCAAGAGAGCTGCTCTGATGAGGACGATGGGCTTCCATCAGCCTTCCAGCTGCC 1080
DB 1021 CCCAAGAGAGCTGCTCTGATGAGGACGATGGGCTTCCATCAGCCTTCCAGCTGCC 1080
QY 1081 CACGAGCTGGGCGACGTTTCAACATGCCCCATGACATGTGAAGTCTGTGAGGAGTG 1140
DB 1081 CACGAGCTGGGCGACGTTTCAACATGCCCCATGACATGTGAAGTCTGTGAGGAGTG 1140
QY 1141 TTTGGGAAGCTCCGAGGCCAACACATGATGTCGCCGACCTCATCCAGTCCAGCGTCC 1200
DB 1141 TTTGGGAAGCTCCGAGGCCAACACATGATGTCGCCGACCTCATCCAGTCCAGCGTCC 1200
QY 1201 AACCCCTGTGAGCTGAGTGTGCTGATCAACCGATATCCAGCGGCGGCGACGCT 1260
DB 1201 AACCCCTGTGAGCTGAGTGTGCTGATCAACCGATATCCAGCGGCGGCGACGCT 1260
QY 1261 GACTGCTCTCGAGCAACCGAGCAAGCCATCTCCCTGCGGAGGATCTGCGGGCGCC 1320
DB 1261 GACTGCTCTCGAGCAACCGAGCAAGCCATCTCCCTGCGGAGGATCTGCGGGCGCC 1320
QY 1321 AGCTACACCTTGAGCGAGGCTGGAGTGGCTTTTGGCGTGGGCTCCAGCCCTGTCT 1380

DB 1321 AGCTACACCTTGAGCGAGGCTGGAGTGGCTTTTGGCGTGGGCTCCAGCCCTGTCT 1380
QY 1381 TACATGCACTACTGACCAAGCTGTGGTGCACCGGGAAGGCCAAGGACAGATGTGTGTC 1440
DB 1381 TACATGCACTACTGACCAAGCTGTGGTGCACCGGGAAGGCCAAGGACAGATGTGTGTC 1440
QY 1441 CAGACCGGCCACTTCCCTGGGCGGATGGCACCAGCTGTGGGAGGGAAGCTGTGCTC 1500
DB 1441 CAGACCGGCCACTTCCCTGGGCGGATGGCACCAGCTGTGGGAGGGAAGCTGTGCTC 1500
QY 1501 AAAGGGCGCTGCTGGAGAGACACCACTTCAACAAGCACAG 1541
DB 1501 AAAGGGCGCTGCTGGAGAGACACCACTTCAACAAGCACAG 1560
QY 1542 -----GCTGATGCT 1551
DB 1561 TCTCCAAAACAACTTATTAAAGGCTACAAATGGCTGCACACTACACAGGTGATGCT 1620
QY 1552 TCTCGGGCAATGGATCCCTATGCGCCCTGCTGCGGCACATGTGGTGGGGGTGCGAG 1611
DB 1621 TCTCGGGCAATGGATCCCTATGCGCCCTGCTGCGGCACATGTGGTGGGGGTGCGAG 1680
QY 1612 CTGGCCAGGAGGAGTGCACCAACCCCTGTCACACGGGGCAAGTACTGCGAGGA 1671
DB 1681 CTGGCCAGGAGGAGTGCACCAACCCCTGTCACACGGGGCAAGTACTGCGAGGA 1740
QY 1672 GTGAGGGTGAATACCGGATCTGCAACCTGGAGCCCTGCCCGAGCTCAGCTCCGGAAAG 1731
DB 1741 GTGAGGGTGAATACCGGATCTGCAACCTGGAGCCCTGCCCGAGCTCAGCTCCGGAAAG 1800
QY 1732 AGCTTCCGGGAGGAGCTGTGAGGCTTTCACGCGCTTCAACGCTACACACAGCAACCGGCTC 1791
DB 1801 AGCTTCCGGGAGGAGCTGTGAGGCTTTCACGCGCTTCAACGCTACACACAGCAACCGGCTC 1860
QY 1792 ACTCTGCGCGTGGCATGGGTGCCCAAGTACTCCGCGCTGTCTCCCGGGCAAGTGAAG 1851
DB 1861 ACTCTGCGCGTGGCATGGGTGCCCAAGTACTCCGCGCTGTCTCCCGGGCAAGTGAAG 1920
QY 1852 CTCATCTCCCGAGCAATGGGACCTGAGCTTCTATGCTGCGGACCCAGGCTGAGAC 1911
DB 1921 CTCATCTCCCGAGCAATGGGACCTGAGCTTCTATGCTGCGGACCCAGGCTGAGAC 1980
QY 1912 GGCAGCTGTGCTCTCTGACTCCACCTCCGCTGTGTCTTCCAAAGCAAGTGCATCAAGCT 1971
DB 1981 GGCAGCTGTGCTCTCTGACTCCACCTCCGCTGTGTCTTCCAAAGCAAGTGCATCAAGCT 2040
QY 1972 GGCTGTGATGGGAACCTGGGCTCCAAAGAGATTCGACAAAGTGTGGGGTGTG*GGGGGA 2031
DB 2041 GGCTGTGATGGGAACCTGGGCTCCAAAGAGATTCGACAAAGTGTGGGGTGTG*GGGGGA 2100
QY 2032 GACATAAGAGCTGCAAGAAGTGAAGTGGGCTTTCACCAAGCCCATGCTGCTGCTCAAT 2091
DB 2101 GACATAAGAGCTGCAAGAAGTGAAGTGGGCTTTCACCAAGCCCATGCTGCTGCTCAAT 2160
QY 2092 TTCGTGTGGCCATCCCGCAGGCGCTCAAGCATGACATCCCGCAGCGGGTTACAAA 2151
DB 2161 TTCGTGTGGCCATCCCGCAGGCGCTCAAGCATGACATCCCGCAGCGGGTTACAAA 2220
QY 2152 GGGCTGATCGGGGATGACAACTACTGGCTCTGAAGAACAGCAAGTACCTGCTC 2211
DB 2221 GGGCTGATCGGGGATGACAACTACTGGCTCTGAAGAACAGCAAGTACCTGCTC 2280
QY 2212 AACGGGCAATTCGTGGTGTGGCGGCTGGAGCGGACCTGGTGGTGAAGGCAAGTCTGCTG 2271
DB 2281 AACGGGCAATTCGTGGTGTGGCGGCTGGAGCGGACCTGGTGGTGAAGGCAAGTCTGCTG 2340
QY 2272 CGGTACAGCGCACGGGACAGCGGTGGAGGCTTCCAGGCTTCCCGGGCCATCTCTGGAG 2331
DB 2341 CGGTACAGCGCACGGGACAGCGGTGGAGGCTTCCAGGCTTCCCGGGCCATCTCTGGAG 2400
QY 2332 CCGCTGACCGTGGAGGCTCTCTCGTGGGGAAGTGAACCGCGCGCGGGTCCCG 2385

Db 2401 CCGCTGACCGTGGAGGTCTCTCCGTGGGAAGATGACACCGCCCGGACCTGC 2454

RESULT 4

US-10-163-316-1

; Sequence 1, Application US/10163316

; Publication No. US20020197703A1

; GENERAL INFORMATION:

; APPLICANT: Kapeller-Libermann, Rosana

; TITLE OF INVENTION: 6552, A Human Matrix Metalloproteinase and Uses

; FILE OF INVENTION: Therefor

; FILE REFERENCE: MPI01-025PIRNM

; CURRENT APPLICATION NUMBER: US/10/163,316

; CURRENT FILING DATE: 2002-06-05

; PRIOR APPLICATION NUMBER: 60/297,863

; PRIOR FILING DATE: 2001-06-13

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 2940

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (472)...(2941)

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(2940)

; OTHER INFORMATION: n = A, T, C or G

US-10-163-316-1

Query Match 80.5%; Score 2296.4; DB 9; Length 2940;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 2379; Conservative 0; Mismatches 6; Indels 69; Gaps 1;

QY 1 ATGCTTTTGTGGCANTCCTAACCCCTGGCTTCGCCGGGGGAACCCGCTGGAGGCTTTGAG 60
DB 472 ATGCTTTTGTGGCANTCCTAACCCCTGGCTTCGCCGGGGGAACCCGCTGGAGGCTTTGAG 531
QY 61 CCAGAGCGGGAGGTAGTCTTCCATCCGACTGGACCCGGGACATTAACGGCCCGCGGTAC 120
DB 532 CCAGAGCGGGAGGTAGTCTTCCATCCGACTGGACCCGGGACATTAACGGCCCGCGGTAC 591
QY 121 TACTGGGGGTCGCCGAGGACTCCGGGGATCAGGAGTCAATTTTCAGATCAAGCATTT 180
DB 592 TACTGGGGGTCGCCGAGGACTCCGGGGATCAGGAGTCAATTTTCAGATCAAGCATTT 651
QY 181 CAGGAGGACTTTTACCTACACCTCAGCCGGATGCTCAGTTCTTGGCTCCGCGCTTCTCC 240
DB 652 CAGGAGGACTTTTACCTACACCTCAGCCGGATGCTCAGTTCTTGGCTCCGCGCTTCTCC 711
QY 241 ACTGAGCATCTGGCGTCCCGCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGCGACGC 300
DB 712 ACTGAGCATCTGGCGTCCCGCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGCGACGC 771
QY 301 TGTCTTATCTGGGGAGCTGAACCGCGAGCTCGTTCGCTGCTGTGAGGCTGTGC 360
DB 772 TGTCTTATCTGGGGAGCTGAACCGCGAGCTCGTTCGCTGCTGTGAGGCTGTGC 831
QY 361 GGGGGGCTCCGCGAGCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCGCTGCC 420
DB 832 GGGGGGCTCCGCGAGCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCGCTGCC 891
QY 421 AATGCTAGCGCGGGGCGAGCGAACAGCCAGGCGCGACACCTTCTCCAGCGCGG 480
DB 892 AATGCTAGCGCGGGGCGAGCGAACAGCCAGGCGCGACACCTTCTCCAGCGCGG 951
QY 481 GGTGTTCCGGCGGGCTTCCGAGAGACCCACCTCTCTGCTGCGGGGTGGGCTCGGGCTG 540
DB 952 GGTGTTCCGGCGGGCTTCCGAGAGACCCACCTCTCTGCTGCGGGGTGGGCTCGGGCTG 1011
QY 541 AACCCCGCATCTACGGGCTTGAGCCCTTACAAAGCGCGGCGGGCTTCGGGGAG 600
|||||

Db 1012 AACCCCGCATCTCTACGGGCGCTTGAGCCCTTACAAAGCGCGGGGGGCTTCGGGGAG 1071
QY 601 AGTCGTAGCGCGCGAGGTCTGGGCGCGCCAAAGGTTTCGTGTCTATATCCCTCGGTACGTG 660
DB 1072 AGTCGTAGCGCGCGAGGTCTGGGCGCGCCAAAGGTTTCGTGTCTATATCCCTCGGTACGTG 1131
QY 661 GAGACGCTGGTGTCTCGCGAGGAGTCAATGTCAAGTTTCCAGCGCGGGA/CTGGAACAT 720
DB 1132 GAGACGCTGGTGTCTCGCGAGGAGTCAATGTCAAGTTTCCAGCGCGGAGCTGGAACAT 1191
QY 721 TATCTGCTGACGCTGCTGGCAACGCGGCGGACTTACCGCATCCAGCATCTCTCAAC 780
DB 1192 TATCTGCTGACGCTGCTGGCAACGCGGCGGACTTACCGCATCCAGCATCTCTCAAC 1251
QY 781 CCATCAACATCGTTGTGTCTCAAGGTGTCTTCTTAGAGATCGTGTACTCGGGCCCCAAG 840
DB 1252 CCATCAACATCGTTGTGTCTCAAGGTGTCTTCTTAGAGATCGTGTACTCGGGCCCCAAG 1311
QY 841 GTCACGGGAATGCGGCGCTGACGCTGGCAACCTTCTGTGCTGGCAGAAAGCTGAAC 900
DB 1312 GTCACGGGAATGCGGCGCTGACGCTGGCAACCTTCTGTGCTGGCAGAAAGCTGAAC 1371
QY 901 AAAGTGAGTGACAAGCACCCCGAGTACTGGGACACTGCGCATCTCTTACCAGGCGAGAC 960
DB 1372 AAAGTGAGTGACAAGCACCCCGAGTACTGGGACACTGCGCATCTCTTACCAGGCGAGAC 1431
QY 961 CTGTGTGGAGCCACCACCTGTGACACCTTGGCATGGGTGTGGTGTGGTGTGGTGTGGTGTGG 1020
DB 1432 CTGTGTGGAGCCACCACCTGTGACACCTTGGCATGGGTGTGGTGTGGTGTGGTGTGGTGTGG 1491
QY 1021 CCCAAGAGAGTGTCTGTCTGATTGAGGAGCATGGGCTTCCATCAGCTTCCACCACTGCC 1080
DB 1492 CCCAAGAGAGTGTGTCTGATTGAGGAGCATGGGCTTCCATCAGCTTCCACCACTGCC 1551
QY 1081 CACGAGCTGGGCGAGGTGTCAACATGCCCCATGACATGTGAAAGTCTGTGAGGAGGTG 1140
DB 1552 CACGAGCTGGGCGAGGTGTCAACATGCCCCATGACATGTGAAAGTCTGTGAGGAGGTG 1611
QY 1141 TTTGGAGACTCCGAGCAACACATGATGTCGCCGAGCTTCCAGCATCCAGATCGACCGTGC 1200
DB 1612 TTTGGAGACTCCGAGCAACACATGATGTCGCCGAGCTTCCAGCATCCAGATCGACCGTGC 1671
QY 1201 AACCCCTGTGAGCTGTCAGTGTGCTCATCATCACGACTTCTTGACAGAGCGGCGACGCT 1260
DB 1672 AACCCCTGTGAGCTGTCAGTGTGCTCATCATCACGACTTCTTGACAGAGCGGCGACGCT 1731
QY 1261 GACTGTCTCTTGACCAACCCAGCAAGCCCATCTCCCTGCCGAGGATCTTCCCGGGCGCC 1320
DB 1732 GACTGTCTCTTGACCAACCCAGCAAGCCCATCTCCCTGCCGAGGATCTTCCCGGGCGCC 1791
QY 1321 AGTACACCTGAGCGAGCTGCGAGCTGGGCTTTTGGGCTGGGCTTCCAGGCTTGTCT 1380
DB 1792 AGTACACCTGAGCGAGCTGCGAGCTGGGCTTTTGGGCTGGGCTTCCAGGCTTGTCT 1851
QY 1381 TACATCAGTACTGCAACCAAGCTGTGTGTCACCGGGAAGGCCAAGGAGACAGATGTTGTGC 1440
DB 1852 TACATCAGTACTGCAACCAAGCTGTGTGTCACCGGGAAGGCCAAGGAGACAGATGTTGTGC 1911
QY 1441 CAGACCCGCGACTTCCCTTGGGCGGATGCGACAGCTGTGGGCGAGGCGAAGCTCTGCTC 1500
DB 1912 CAGACCCGCGACTTCCCTTGGGCGGATGCGACAGCTGTGGGCGAGGCGAAGCTCTGCTC 1971
QY 1501 AAAGGGGCTGCGTGAGAGACACACCTTCAACAGCACAG----- 1541
DB 1972 AAAGGGGCTGCGTGAGAGACACACCTTCAACAGCACAGACACAGCGCTCCCGACTGACATCAT 2031
QY 1542 -----GGTGGATGCT 1551
DB 2032 TCTCCAAAACACTCTTATTAAGGCTACCAATAGGCTTGCACACTACACAGCTGGATGCT 2091
QY 1552 TCTTGGGCCAAATGGGATCCCTATGGCCCTTGTGCGGCACATGTGGTGGGGCGGTGAC 1611
DB 2092 TCCTGGGCCAAATGGGATCCCTATGGCCCTTGTGCGGCACATGTGGTGGGGCGGTGAC 2151
|||||

QY 1612 CTGCCAGGAGGAGTGCACCAACCCACCCCTGCCAACGGGGGAGTACTCGGAGGA 1671
 Db 2152 CTGCCAGGAGGAGTGCACCAACCCACCCCTGCCAACGGGGGAGTACTCGGAGGA 2211
 QY 1672 GTGAGGTGAATACCGATCCTGCAACCTGGAGCCCTGCCAGCTCAGCTCCGGAAG 1731
 Db 2212 GTGAGGTGAATACCGATCCTGCAACCTGGAGCCCTGCCAGCTCAGCTCCGGAAG 2271
 QY 1732 AGCTTCGGGAGGAGGAGTGTGAGGCTTTCAACGGCTACACACAGCACCGGCTC 1791
 Db 2272 AGCTTCGGGAGGAGGAGTGTGAGGCTTTCAACGGCTACACACAGCACCGGCTC 2331
 QY 1792 ACTCTCGCCGTGGCATGGGTGCCAAAGTACTCCGGCGTGTCTCCCGGACAGTCAAG 1851
 Db 2332 ACTCTCGCCGTGGCATGGGTGCCAAAGTACTCCGGCGTGTCTCCCGGACAGTCAAG 2391
 QY 1852 CTATCTGCCAGGCAATGGCACTGGCTACTTCTATGTGCTGGCACCCAGGTGGTGC 1911
 Db 2392 CTATCTGCCAGGCAATGGCACTGGCTACTTCTATGTGCTGGCACCCAGGTGGTGC 2451
 QY 1912 GGCACGCTGTGCTCTCTGACTCCACTCCGCTGTGTCCAGGCAAGTGCATCAAGCT 1971
 Db 2452 GGCACGCTGTGCTCTCTGACTCCACTCCGCTGTGTCCAGGCAAGTGCATCAAGCT 2511
 QY 1972 GGCCTGTGATGGGAACCTGGGCTCCAAAGAGAGATTCCGCAAGTGTGGGGGA 2031
 Db 2512 GGCCTGTGATGGGAACCTGGGCTCCAAAGAGAGATTCCGCAAGTGTGGGGGA 2571
 QY 2032 GACAATAAGAGCTGCAAGAGGTGACTGACTCTTCCACCAAGCCATGATGGCTACAT 2091
 Db 2572 GACAATAAGAGCTGCAAGAGGTGACTGACTCTTCCACCAAGCCATGATGGCTACAT 2631
 QY 2092 TTCGTGGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCGGCGCGGTTTACAAA 2151
 Db 2632 TTCGTGGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCGGCGCGGTTTACAAA 2691
 QY 2152 GGCTCATCGGGATGACAACTACCTGGCTCTGAAGACAGCCAAAGTACCTGCTC 2211
 Db 2692 GGCTCATCGGGATGACAACTACCTGGCTCTGAAGACAGCCAAAGTACCTGCTC 2271
 QY 2212 AACGGGCAATTCGTGTGTGTCGGCGGTGGAGCGGACCTGGTGTGAAGGGAGTCTGTC 2271
 Db 2752 AACGGGCAATTCGTGTGTGTCGGCGGTGGAGCGGACCTGGTGTGAAGGGAGTCTGTC 2811
 QY 2272 CGGTACAGCGGACAGGGGACAGCGGTGGAGAGCCCTGCAGGCTTCCCGCCATCCTGGAG 2331
 Db 2812 CGGTACAGCGGACAGGGGACAGCGGTGGAGAGCCCTGCAGGCTTCCCGCCATCCTGGAG 2871
 QY 2332 CCGCTGACCGTGGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGTCGGC 2385
 Db 2872 CCGCTGACCGTGGAGTCTCTCCGTGGGGAAGATGACACCGCCCGGACCTGC 2925

RESULT 5

US-09-965-631-5
 ; Sequence 5, Application US/09965631
 ; Patent No. US20020115842A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Friddle, Carl Johan
 ; APPLICANT: Hilbun, Erin
 ; TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucleotides Encodin
 ; FILE REFERENCE: LEX-0241-USA
 ; CURRENT APPLICATION NUMBER: US/09/965,631
 ; CURRENT FILING DATE: 2001-09-27
 ; PRIOR FILING DATE: 2000-09-29
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5
 ; LENGTH: 1104
 ; TYPE: DNA
 ; ORGANISM: homo sapiens

US-09-965-631-5

Query Match 38.1%; Score 1088; DB 10; Length 1104;
 Best Local Similarity 99.5%; Pred. No. 4e-281;
 Matches 1091; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGCTTTTGTGGGATCTTAACCCCTGGCTTTCGCGGGGAAACGCTGGAGGCTTTGAG 60
 Db 1 ATGCTTTTGTGGGATCTTAACCCCTGGCTTTCGCGGGGAAACGCTGGAGGCTTTGAG 60
 QY 61 CCAGAGCGGGAGGTAGTCTTCCCATCGACTGGACCGGACATTAACGGCGCGCGCTAC 120
 Db 61 CCAGAGCGGGAGGTAGTCTTCCCATCGACTGGACCGGACATTAACGGCGCGCGCTAC 120
 QY 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCATTTTTCAGATCACAGCATTT 180
 Db 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCATTTTTCAGATCACAGCATTT 180
 QY 181 CAGGAGGACTTTTACCTACACTGACGCGCGGATGCTCAGTTCTTGGCTCCGCTTCTCC 240
 Db 181 CAGGAGGACTTTTACCTACACTGACGCGCGGATGCTCAGTTCTTGGCTCCGCTTCTCC 240
 QY 241 ACTGAGCATCTGGGCTCCCGCTCCAGGGGCTCCAGGGGCTCTTCCAGACTCTCGACGC 300
 Db 241 ACTGAGCATCTGGGCTCCCGCTCCAGGGGCTCCAGGGGCTCTTCCAGACTCTCGACGC 300
 QY 301 TGTCTTATTCTGGGAGGTGAACCGCGAGCGGACTCGTTGCTGTGTGAGCTGTGC 360
 Db 301 TGTCTTATTCTGGGAGGTGAACCGCGAGCGGACTCGTTGCTGTGTGAGCTGTGC 360
 QY 361 GGGGGCTCCCGGAGCCTTTGGCTACCGAGGCGCGGATGTCATTTAGCCGCTGCC 420
 Db 361 GGGGGCTCCCGGAGCCTTTGGCTACCGAGGCGCGGATGTCATTTAGCCGCTGCC 420
 QY 421 AATGCTAGCGCGCGGGCGGCGCAACAGCAGCGCGCACACCTTCTCCAGCGCGG 480
 Db 421 AATGCTAGCGCGCGGGCGGCGCAACAGCAGCGCGCACACCTTCTCCAGCGCGG 480
 QY 481 GGTGTTCGGGCGCGGCTTCGGAGAGCCACCTCTCGCTCGGGGTGGCTTCGCGTGG 540
 Db 481 GGTGTTCGGGCGCGGCTTCGGAGAGCCACCTCTCGCTCGGGGTGGCTTCGCGTGG 540
 QY 541 AACCCCGCCATCTTACGCGGCTTGAGCCCTTACAAGCGCGGGCGGGGCTTCGGGGAG 600
 Db 541 AACCCCGCCATCTTACGCGGCTTGAGCCCTTACAAGCGCGGGCGGGGCTTCGGGGAG 600
 QY 601 AGTGTAGCGCGGCGGCTTCGGGCGCGCAAGGCTTTCGTGTATTCGCGGCTAGCTG 660
 Db 601 AGTGTAGCGCGGCGGCTTCGGGCGCGCAAGGCTTTCGTGTATTCGCGGCTAGCTG 660
 QY 661 GAGACGCTGTGTGTCGGGAGAGTCAATGGTCAAGTTCCACGCGCGGAGCTTGGACAT 720
 Db 661 GAGACGCTGTGTGTCGGGAGAGTCAATGGTCAAGTTCCACGCGCGGAGCTTGGACAT 720
 QY 721 TATCTGTGAGCTGTGGCAAGCGGCGGCTTACCGCCATCCAGCATCTCCCACTCAAC 780
 Db 721 TATCTGTGAGCTGTGGCAAGCGGCGGCTTACCGCCATCCAGCATCTCCCACTCAAC 780
 QY 781 CCATCAACATCGTTGTGGTCAAGGCTCTTCTTAGAGATCGTACTCCGGGCGGCAAG 840
 Db 781 CCATCAACATCGTTGTGGTCAAGGCTCTTCTTAGAGATCGTACTCCGGGCGGCAAG 840
 QY 841 GTCACCGGCAATCGGGCCCTGACGCTGGCAACTTCTGTGCTGGCAAGAGAGCTGAAC 900
 Db 841 GTCACCGGCAATCGGGCCCTGACGCTGGCAACTTCTGTGCTGGCAAGAGAGCTGAAC 900
 QY 901 AAGTGTGAGTGCACAGCACCCGAGTACTGGGACACTGCGCATCTCTTCCAGGCGAGGAC 960
 Db 901 AAGTGTGAGTGCACAGCACCCGAGTACTGGGACACTGCGCATCTCTTCCAGGCGAGGAC 960
 QY 961 CTGTGTGAGGCGCACCTCTGTACACCTCGGCGATGCTGATGGTACCATTGTGTGAC 1020
 Db 961 CTGTGTGAGGCGCACCTCTGTACACCTCGGCGATGCTGATGGTACCATTGTGTGAC 1020

QY 1021 CCCAAGAGAGCTGCTCTGTCTATTTGAGGACGATGGCTTCCATCAGCCTTACCAGTCC 1080
Db 1021 CCCAAGAGAGCTGCTCTGTCTATTTGAGGACGATGGCTTCCATCAGCCTTACCAGTCC 1080

QY 1081 CACGAGCTGGCCACG 1096
Db 1081 CACGAGCTGGGTRAGG 1096

RESULT 6

US-09-965-631-1
; Sequence 1, Application us/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucleotides Encodin
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 966
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-965-631-1

Query Match . 33.5% Score 955.8; DB 10; Length 966;
Best Local Similarity 99.88; Pred. No. 8.4e-246;
Matches 957; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCTTTTGTGGGATCTTAACCTGGCTTTCGCGGGGGAACCGCTGGAGGCTTTGAG 60
Db 1 ATGCTTTTGTGGGATCTTAACCTGGCTTTCGCGGGGGAACCGCTGGAGGCTTTGAG 60

QY 61 CCAGAGCGGGAGGTAGTCTTCCATCCGACTGCGACCGGACATTAACGCCCGCGCTAC 120
Db 61 CCAGAGCGGGAGGTAGTCTTCCATCCGACTGCGACCGGACATTAACGCCCGCGCTAC 120

QY 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGTCAATTTTCAGATCAACAGATTT 180
Db 121 TACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGTCAATTTTCAGATCAACAGATTT 180

QY 181 CAGGAGGACTTTTACCTACACTGACCGCGGATGCTCAGTTCTTGGCTCCCGCTTCTCC 240
Db 181 CAGGAGGACTTTTACCTACACTGACCGCGGATGCTCAGTTCTTGGCTCCCGCTTCTCC 240

QY 241 ACTGAGCATCTGGGCTCCCGCTCCAGGGGCTCAGCGGGGCTCTTCAGACTCGGACGC 300
Db 241 ACTGAGCATCTGGGCTCCCGCTCCAGGGGCTCAGCGGGGCTCTTCAGACTCGGACGC 300

QY 301 TGTCTTATCTTGGGAGCTGAACCGCGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
Db 301 TGTCTTATCTTGGGAGCTGAACCGCGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360

QY 361 GGGGGCTCCCGGAGCTTTGGCTACCGAGCGCGGAGTATGTCATTAGCCCGCTGCC 420
Db 361 GGGGGCTCCCGGAGCTTTGGCTACCGAGCGCGGAGTATGTCATTAGCCCGCTGCC 420

QY 421 AATGCTAGCGCGCGCGGCGAGCGCAACAGCCAGGCGGCACACCTTCTCCAGCGCGG 480
Db 421 AATGCTAGCGCGCGCGGCGAGCGCAACAGCCAGGCGGCACACCTTCTCCAGCGCGG 480

QY 481 GGTGTTCCGGCGGGCTTTCGGAGACCCACCTCTCGCTGGGGGTGGCTCGGGCTGG 540
Db 481 GGTGTTCCGGCGGGCTTTCGGAGACCCACCTCTCGCTGGGGGTGGCTCGGGCTGG 540

QY 541 AACCCCGCATCTTACGGGCGCTTACAGCTTACAAAGCGCGCGGGCGGCTTCCGGGAG 600
Db 541 AACCCCGCATCTTACGGGCGCTTACAGCTTACAAAGCGCGCGGGCGGCTTCCGGGAG 600

Db 541 AACCCCGCATCTTACGGGCGCTTACAGCTTACAAAGCGCGCGGGCGGCTTCCGGGAG 600
QY 601 AGTCGTAGCGCGCGAGTCTTGGCGCGCCCAAGGTTTCGTGTCTATCCCGCGGTACGTG 660
Db 601 AGTCGTAGCGCGCGAGTCTTGGCGCGCCCAAGGTTTCGTGTCTATCCCGCGGTACGTG 660
QY 661 GAGACGCTGGTGGTCCGCGGAGGTCATATGTCAAGTTTCCACGGCGGAGCTGGAACAT 720
Db 661 GAGACGCTGGTGGTCCGCGGAGGTCATATGTCAAGTTTCCACGGCGGAGCTGGAACAT 720
QY 721 TATCTGCTGACGCTGCTGCGACGCGGCGGACTTACCGCCATCCAGCATCTCTCAAC 780
Db 721 TATCTGCTGACGCTGCTGCGACGCGGCGGACTTACCGCCATCCAGCATCTCTCAAC 780
QY 781 CCCATCAACATCGTTGTGTCAAGTGTCTTCTTAGAGATCGTGACTCGGGGCCAAG 840
Db 781 CCCATCAACATCGTTGTGTCAAGTGTCTTCTTAGAGATCGTGACTCGGGGCCAAG 840
QY 841 GTCACCGGCAATGCGGCGCTGAGCTGGCAACTTCTGTGCTGGCAGCAAGAGCTGAAC 900
Db 841 GTCACCGGCAATGCGGCGCTGAGCTGGCAACTTCTGTGCTGGCAGCAAGAGCTGAAC 900
QY 901 AAAGTGAGTGACAAGCACCAGTACTGGGACACTGCGCATCTCTTACCAGGTCAGGA 959
Db 901 AAAGTGAGTGACAAGCACCAGTACTGGGACACTGCGCATCTCTTACCAGGTCAGGA 959

RESULT 7

US-10-097-597-13
; Sequence 13, Application US/10097597
; Publication No. US20030022352A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inouchi, Eiichi
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 2184
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1)..(2184)
US-10-097-597-13

Query Match 22.3%; Score 636; DB 9; Length 2184;
Best Local Similarity 60.7%; Pred. No. 2.8e-160;
Matches 1098; Conservative 0; Mismatches 700; Indels 12; Gaps 3;

QY 631 AAGGTTTCTGCTCTATCCCGCGGTAGCTGGAGACGCTGGTGTGCGGACGAGTCAATG 690
Db 31 AAGGATTTTCTGCTGACGCGGCTTATGTGGAACCATGCTCTAGCTACCAAGTCCATG 90

QY 691 GTCAAGTTTCCAGCGCGGACCTGGAACATTTATCTGTGACGCTGTGGCAACGCGGCG 750
Db 91 GCGGACTTCCAGCGGCGGCTCTAAGCATTTACCTTCTTAACCTGTTCTCGGTGGCAGCC 150

QY 751 CGAATCTACCGCCCATCCAGCATCTCAACCCCATCAACATCGTTGTGTCAGGTGCTG 810

[illegible]

Db 1288 GCCGGCGCTCTCGCCAAAGGACAGGTCGAAGCTCACTGTGAAGCCAAAGGCAATTGGCTAC 1287
 QY 1882 TTCTATGTGCTGGCACCACCAAGGTGGTGGAGCGGCAGCGCTGTGCTCTCTCTGAGTCCACGCTCC 1941
 Db 1288 TTTTTCGCTCTACAGCCCAAGGTTGTAGATGGCACTCCCTGTAGTCTCAGACTCTACCTCT 1347
 QY 1942 GTCTGTGTCCAAAGCAAGTCGATCATAAGGCTGTGATGGGAACCTGGGTGCCAAGAAAG 2001
 Db 1348 GTCTGTGTGCAAGGGCAGTGTGTGAAGAGCTGGCTGTGATGCGCATATAGACTCCGAAAG 1407
 QY 2002 AGATTTCACAAGTGTGGGGTGTCTGGGGGAGACAAATAAGAGGCTGCAAGAAGGTGACTGGA 2061
 Db 1408 AAGTTTGATAGTGTGCGTTTCTGGAGGAACGGTTCACATGCAAGAAGATGTCAGGA 1467
 QY 2062 CTCTTCACCAAGCCCATGCAATGCTACAATTCTGTGTGGGCCATCCCGCAGGCGCCTCA 2121
 Db 1468 ATAGTCACTAGTACAGACCTGGGTATCATGACATTTGTCAAAATTCCTGTCTGGAGCCACC 1527
 QY 2122 AGCATCAGACATCCGCCAGCGGTTACAAGAGGCTGATCGGGGATGACAACCTACCTGCT 2181
 Db 1528 AACATTGAAGTGAACATCGGAATCAAGGGGTCCTCAGAAACAATGGCAGCTTTCTGGCT 1587
 QY 2182 CTGAAGAACCAAGCAAGTACCTTGCTCAACGGGCAATTTCTGTGTGGGCCATCCCGCAGGCTGGAG 2241
 Db 1588 ATTAGACCGCTGATGTACCTATATTTCTGAATGGTAAACTTCACTCTGTGCACACTAGAG 1647
 QY 2242 CGGGACCTGTGTGTAAGGCGAGTCTGCTGCGGTACAGGGGCACGGGCACAGCGGTGGAG 2301
 Db 1648 CAAGACCTCACTACAAGAGTACTGCTTTAAGGTACAGTGGTTCTCGGGCTCGGCTGGAA 1707
 QY 2302 AGCGTCGAGGCTTCCGGGGCCATCTCTGGAGCGCTGACCGTGGAGGTCTCTCGCTGGGG 2361
 Db 1708 AGAATCCGCAGCTTTAGTCCACTCAAGAAGACCTTTAACCATCCAGGTCTTTATGTTAGGC 1767
 QY 2362 AAGATGACACCGCCCGGGTCCGCTACTCTTCTATCTGCCCCAAGAGCCTCTGGGGAGAC 2421
 Db 1768 CATGCTCTCGAGCCCAAAATTAATTCACCTACTTTATGAAGAAGACAGAGTCAATC 1827
 QY 2422 AAGTCCTCTC 2431
 Db 1828 AACGCCATTC 1837
 RESULT 8
 US-10-097-580-13
 ; Sequence 13, Application US/10097580
 ; Publication No. US20030032168A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hirose, Kunitaka
 ; APPLICANT: Inoguchi, Biiji
 ; APPLICANT: Hakozaeki, Michinori
 ; APPLICANT: Ishioaka, Keiko
 ; APPLICANT: Ishida, Yukako
 ; APPLICANT: Matsushima, Kouji
 ; APPLICANT: Kuno, Kouji
 ; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharm
 ; TITLE OF INVENTION: composition and method of immunologically analyzing
 ; FILE REFERENCE: Q57092
 ; CURRENT APPLICATION NUMBER: US/10/097,580
 ; CURRENT FILING DATE: 2002-03-15
 ; PRIOR APPLICATION NUMBER: 09/445,023
 ; PRIOR FILING DATE: 1999-12-03
 ; PRIOR APPLICATION NUMBER: JP 9-160422
 ; PRIOR FILING DATE: 1997-06-03
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 13
 ; LENGTH: 2184
 ; TYPE: DNA
 ; ORGANISM: Mus sp.
 ; FEATURE:
 ; NAME/KEY: exon
 ; LOCATION: (1)..(2184)

```

RESULT 8
US/10/580-13
; Sequence 13, Application US/10097580
; Publication No. US20030032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakoizaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein
; TITLE OF INVENTION: composition and method
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 2184
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1)..(2184)

```


US-10-097-580-13

```

Query Match      22.3%; Score 636; DB 9; Length 2184;
Best Local Similarity 50.7%; Pred. No. 2.8e-160;
Matches 1098; Conservative 0; Mismatches 700; Indels 12; Gaps 3;

QY 631 AAGCGTTTCGTGCTATCCCGCGTACGTGGAGACGCTGGTGTCTCGCGACGAGTCAATG 690
DB 31 AAGCGATTTGTCTCCAGCCCGCTTATGTGAAACCATGCTCGTAGCTGACCACTCATG 90

QY 691 GTCAGTTTCACGGCGGGACCTGGAACATATCTGCTGACGCTGCTGGCAACGGCGGG 750
DB 91 GCCGACTTCCACGCGCAGCGTCTAAAGCATTTACCTTCTTACCCCTGTCTCGTGGCAGCC 150

QY 751 CGACTCTACCGCCATCCACGATCCCTCAACCCCATCAACATCTGTGTGTCAGGTGCTG 810
DB 151 AGGTTTACAGCATCCCGACATTTAGGAATTCATTTAGCTGTGTGGTGGTGAAGATCTTG 210

QY 811 CTTCTTAGACTCTGTACTCCGGGCCCAAGGTCAACGGCAATCGGCCCTGACGCTGCGC 870
DB 211 GTCATATACGAGGAGCAGAGGACCAAGAGTTTACCTCCAATGACGCTCTCACCTTCGG 270

QY 871 AACTTCTGTGCTGGAGAGAGCTGAACAAAGTGAAGTCAACAGCACCCGAGTACTGG 930
DB 271 AATTCTGCACTGGCAGAACACACACAGCCCGCAGTACCGGGATCCAGAGCACTAT 330

QY 931 GACACTGCACTCTCTTACACGAGCAGGAGCTGTGTGGAGCCACCACTCTGTGACACCTG 990
DB 331 GACACTGCAATCTGTTTACACAGACAGATTTATGTGGCTCCACAGCTGTGACACTCTC 390

QY 991 GGCATGCTGATGTGGGTACCATGTGTGACCCCAAGAGAGTGTCTGTCTGATTTAGGAG 1050
DB 391 GGAATGGCAGATGTGGAACCGTATGTGACCCCAAGAGAGTGTCTGATTTAGAGAT 450

QY 1051 GATGGCTTCCATCAGCTTCCACACTGCCAGAGCTGGCCAGCTGTTTCAACATGCC 1110
DB 451 GATGGTTGCAAGTGTCTTCCACAGCCCATGAATGGGCCATGTGTTTACATGCCG 510

QY 1111 CATGACATGTGAAGTCTGTGAGAGGTGTTTGGAGAGTCCGAGCAACCACTGATG 1170
DB 511 CACGATGATGTAAGCACTGTGCCAGTGTGAATGGTGTGAGTGGCGATTTCTCATCTGATG 570

QY 1171 TCCCGGACCTTCATCCAGATCGACCTGTCCACACCCCTGGTTCAGCTGCTGCTGCCATC 1230
DB 571 GCCTCGATGCTCTCCAGCTTAGACCATAGCCAGCCCTGGTCACTTCCAGTGGCTACATG 630

QY 1231 ATCACCAGATTTCTTGGCAGCGGCGCGGTGACTGCTCTGAGCAACCAACGCAAGCCC 1290
DB 631 GTCAGTCTTCCCTAGATATATGAGACAGGGGATGTTTGTGACACAGCCCGCAGATCCA 690

QY 1291 ATCTCCCTGCGGAGGATCTGCCGGGCGCCAGCTTACACCTTGACCCAGCAGTGGAGCTG 1350
DB 691 ATCAAGCTCCCTTCTGATCTTCCCGGTACCTTGTACGATGCCAACCAGCTGTGAGTTT 750

QY 1351 GCTTTTGGCGTGGCTCCAACCCCTGTCTTACATGACAGTAC ---TGCACCAAGCTGTGG 1407
DB 751 ACATTCGGAGAGGAATCCAGCACTGCCCCTGTATGACGACGACATGTACTACCCCTGTGG 810

QY 1408 TGCACCGGAAGGCGCAAGGGACAGATGGTGTGCCAGACCCCGCCACTTCCCTCGGCGCAT 1467
DB 811 TGCAGTGCACCTCGGTGGCTTACTGTTGTCACCAACAAACACTTCCCTTGGCGAGAT 870

QY 1468 GGCACAGCTGTGGCGAGGGAAGCTGTGCTCAAGAGGGGCTCGGTGGAGAGACACAAC 1527
DB 871 GGCACAGCTGTGGAGAGGGAAGTGGTGTGTGAGTGGCAAGTGGTGAACAGACAGAC 930

QY 1528 CTCACAGACAGCAG ---GGTGGATGTTCTTGGCCCAATGGATCCCTATGGCCCC 1591
DB 931 ATGAAGCATTTTGGCTACTCTCTGTTGATGGAAGCTGGGACCATGGGACCGTGGGGAGAC 990

QY 1592 TGCTGGCCACATGTGGTGGGGGCTGACGCTGGCCAGGAGGAGTGCACCAACCCCAACC 1641
DB 991 TGCTCAAGAACCTGTGGTGGTGGAGTTCAATACAAATGAGAGATGTGACAAACCCAGTC 1050

```

```

QY 1642 CQTGCAACGGGGCAAGTACTGCGAGGAGTGAAGGTGAAATACCGATCTGCAACCTG 1701
DB 1051 CCAAGAAACGGGAGGAGTACTGTGAAGCAAAACAGTCCGCTACAGTCTCTTAACATC 1110

QY 1702 GAGCCTCGCCCGCAGCTCAGCCTCCCGGAAAGAGCTTCCGGGAGGAGCAGTGTGAGGCTTTC 1761
DB 1111 GAGGACTCTCCAGACAATAA ---CGGAAAACGTTTCAGAGAGGAGCAGTGTGAGGCGGAC 1167

QY 1762 AAGCGCTACAACACACAGCACCAACCGGCTCACTCTCGCGCTGGATGGTGGCTCCCAAGTAC 1821
DB 1168 AATGAGTTTTTCAAAGCTTCTTTGGGAATGAGCCCATCTGTAGAGTGACACCCCAAGTAC 1227

QY 1822 TCCGCGCTGTCTCCCGGACAAAGTGCAGAGTCACTPCTCCGAGCCAAATGGCACTGGCTAC 1881
DB 1228 GCGCGCTCTCCCAAGGACAGAGTGCAGAGTCACTCTCTGAAAGCCAAAGGCAATTTGGCTAC 1287

QY 1882 TTCTATGTGCTGGCACCACCAAGGTGGTGGAGCGGACGCTGTCTCTCTGACTCCACCTCC 1941
DB 1288 TTTTTCGCTTTTACAGCCCAAGGTTGTAGATGCACTCCCTGTAGTCCAGACTCTACCTCT 1347

QY 1942 GTCTGTGTCCAAGGCAAGTGCATCAAGGCTGGCTGTGATGGGAACTGGGCTCCCAAGAG 2001
DB 1348 GTCTGTGTGCAAGGCGAGTGTGTGAAGCTGGCTGTGATCGCATCATAGACTCCCAAGAG 1407

QY 2002 AGATTGCAAGTGTGGGTGTGTGGGGAGAGACAATAAGAGCTGCAAGAGGTGACTGGA 2061
DB 1408 AAGTTTGTATAAGTGTGGGCTTTTGTGGAGGAAACGGTTCCACATGCAAGAGATGTGACGA 1467

QY 2062 CTCTTCACCAACGCCATGCTGCTACAAATTTCTGTGGTGGCCATCCCGCAGGCGCTCA 2121
DB 1468 ATAGTCACTAGTACAAGACCTGGGTATCATGACATTTGCACAAATTCCTGTGGAGCCACC 1527

QY 2122 AGCATGCACTCCGCGAGCGGTTTACAAAGGCTGTATGGGGATGACAACTTACCTGGCT 2181
DB 1528 AACATTGAGTGAACATCGGAATCAAAGGGGTCCAGAAACAAATGGCAGCTTTCTGGCT 1587

QY 2182 CTGAGAACAGCCAGGCAAGTACCTGCTCAACGGGCTTTCTGTGGTGGCGGTGGAG 2241
DB 1588 APTAGAGCGCTGATGGTACCTATATCTGAATGGAACCTTCACCTGTCCACACTAGAG 1647

QY 2242 CGGAGCTGTGTGTAAGGCGAGTCTGCTGCGGTACAGCGGCAGCGGCACAGCGTGGAG 2301
DB 1648 CAAGACTCACCTACAAAGGTACTGTCTTAAGTACAGTGGTCTCTCGCTCGCTGGAA 1707

QY 2302 AGCCTGACGCTTCCCGGCCCATCTCTGGAGCGCTGACCGTGGAGGCTCTCTCCGTGGG 2361
DB 1708 AGAATCCGCGAGCTTTAGTCCACTCAAAAGAACCCCTTAACCATCCAGGTTCTTATGTTAGGC 1767

QY 2362 AAGATGACACCCCGCGGCTCGCTACTCTCTTCTATCTGCCCAAGAGCCTCGGGAGGAC 2421
DB 1768 CATGCTCTCCGACCCCAAAATTAATTTACCTACTTTATGAAGAGACAGAGTCATTC 1827

QY 2422 AAGTCCCTCTC 2431
DB 1828 AAGCCCATTC 1837

```

RESULT 9
 US-09-445-023A-13
 ; Sequence 13, Application US/09445023A
 ; Patent No. US20020119167A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hirose, Kunitaka
 ; APPLICANT: Inoguchi, Eiichi
 ; APPLICANT: Hakoizaki, Michinori
 ; APPLICANT: Ishikawa, Keiko
 ; APPLICANT: Ishida, Yukako
 ; APPLICANT: Matsushima, Kouji
 ; APPLICANT: Kuno, Kouji
 ; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
 ; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA
 ; FILE REFERENCE: Q57092

Sequence 2, Application US/10097597
Publication No. US20030022352A1

GENERAL INFORMATION:

APPLICANT: Hirose, Kunitaka
APPLICANT: Inoguchi, Elji
APPLICANT: Hakozaaki, Michinori
APPLICANT: Ishioaka, Keiko
APPLICANT: Ishida, Yukako
APPLICANT: Matsushima, Kouji
APPLICANT: Kuno, Kouji

TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,

TITLE OF INVENTION: pharmaceutical

TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS

FILE REFERENCE: Q57092

CURRENT APPLICATION NUMBER: US/10/097,597

CURRENT FILING DATE: 2002-03-15

PRIOR APPLICATION NUMBER: 09/445,023

PRIOR FILING DATE: 1999-12-03

PRIOR APPLICATION NUMBER: JP 9-160422

PRIOR FILING DATE: 1997-06-03

NUMBER OF SEQ ID NOS: 14

SOFTWARE: PatentIn version 3.0

SEQ ID NO 2

LENGTH: 2184

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: exon

LOCATION: (1)..(2184)

US-10-097-597-2

Query Match 22.0%; Score 628.8; DB 9; Length 2184;

Best Local Similarity 60.9%; Pred. No. 2.4e-158;

Matches 1080; Conservative 0; Mismatches 682; Indels 12; Gaps 3;

QY 631 AAGGTTTCGTCTATCCCGCGGTAGCTGGAGACGCTGGTGTCTCGGGACGAGTCAATG 690
DB 31 AAGCGATTGTGTCCAGCCCGGTATGTGGAAACCATGCTTGTGGCAGACCAGTGCATG 90
QY 691 GTCAAGTTCACCGCGCGGACCTTGAACATATCTGTGACGCTGCTGGCAACGCGGCG 750
DB 91 GCAGAAATCCACGCGAGTGTCTAAAGCATTAACCTTCTCAGTGTGTTTCGGTGGCAGC 150
QY 751 CGACTACCGCCATCCAGCATCTCAACCCCATCAACATCTGTTGGTCAAGTGTCTG 810
DB 151 AGATTGTACAAACACCCAGCATCTGTAATTCAGTTAGCTGTGTGGTGAAGATCTTG 210
QY 811 CTTCCTTAGAGATCGTGACTCCGGGCCCAAGTTCACCGCAATCGGCCCTGACGCTGCGC 870
DB 211 GTCATCCAGATGAACAGAGGGCCGGAAGTGAACCTCAATGCTGCCCTCACTCTGGG 270
QY 871 AACTTCTGTGCTGGCAGAAAGCTGAACAAAGTGAAGTGAACAGCAACCCCGAGTACTGG 930
DB 271 AACTTTTGAACGTGGCAGAAAGCAGCAACCCAGTACCCGGGATGAGAGCACTAT 330
QY 931 GACACTGCCATCTTTCACCGAGGAGCTGTGGAGCCAGCCAGCTGTGACACCCCTG 990
DB 331 GACACGAATCTTTTTCACGAGACAGACTTGTGGTGGTCCAGAGACATGTGATACTCT 390
QY 991 GGCATGCTGTATGGGTGACATGTGACCCCAAGAGAGCTGCTGTGCTATTGAGGAC 1050
DB 391 GGGATGCTGTATGGAACTGTGTGATCCGAGCAGAGAGCTGCTCCGCTCATTAGAGAT 450
QY 1051 GATGGGCTTCATGACCTTCAACATGCCCCAGAGCTGGGCCAGCTGTTTCAACATGCC 1110
DB 451 GATGGTTTACAAGCTGCTTCAACAGCCCATGAATAGGCCAGCTGTTTAAACATGCCA 510
QY 1111 CATGACATGTGAAGTCTGTGAGGAGTGTGTTGGNAGCTCCGAGCCACCATGATG 1170
DB 511 CATGATGATGCAAGAGCTGTGCGAGCTTAAATGGTGTGAACAGGATTCACCATGATG 570
QY 1171 TCCCGGACCTCATCCAGATCGACCGTGGCCAAACCCCTGGTGCAGCTGCGATCGCATC 1230

DB 571 GCGTCAATGCTTCCAACTGGACACAGCCAGCCCTTGGTCTCTCTTGCAGTGCCTACATG 630
QY 1231 ATCACCGACTTCTTGGACAGCGGCGACGCTGACTGCTCTCTTGGACACCCAGCAAGCCC 1290
DB 631 ATTACATCATTTCTGGATAATGGTTCATGGGAATGTTTGTATGGACAAAGCTTCAGATCCC 690
QY 1291 ATCTCCCTGCGCGAGGAGTCTCCCGGGCGCAGCTACACCTTGAGCCAGCAGCTGGAGCTG 1350
DB 691 ATACAGCTCCCGAGGAGTCTCCCTGGCACCTTGTACGATGCAACCCGCGAGTGCAGTTT 750
QY 1351 GCTTTTGGCTGGGCTCCAAAGCCCTGCTTACATGCAATGACAGTAC---TGCAACCAAGCTTGG 1407
DB 751 ACATTTGGGAGGAGTCCAAACACTGCCCCCATGAGCCAGCAGATGAGACACTTGTGG 810
QY 1408 TGCACCGGGAAGGCAAGGAGCAGATGCTGCCAGACCCCGCCACTTCCCTGGGCCGAT 1467
DB 811 TGTACCGGACCTCTGGTGGGCTGCTGTGTCAAAACCAACACACTTCCCTGTGGCGGAT 870
QY 1468 GGCACACAGCTGTGGCGAGGGAAGCTCTGCTCTCAAGGGGCTGCTGGAGAGACACAAAC 1527
DB 871 GGCACACAGCTGTGGAGAAAGGAATGCTATCAACGCGCAAGTGTGTGAACAAACCGAC 930
QY 1528 CTCACAAAGCACAGG-----GTGGATGTTCTTGGGCAAAATGGGATCCCTATGGCCCC 1581
DB 931 AGGAAGCATTTTGTATACCCCTTTTCATGGAAGCTGGGAGCCATGGGAGCCGTTGGGAGAC 990
QY 1582 TGCTCGGCACATGTGTGGGGGCTGCAGCTGGCCAGAGGAGCAGTGCACCAACCCACCC 1641
DB 991 TGTTCGAGAACGCTGGGTGGAGAGTCCAGTACAGGATGAGGGAATGTGACAAACCCAGTC 1050
QY 1642 CCGTCCACGCGGGGCAAGTACTGCGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1701
DB 1051 CCAAGAAATGGAGGAGTACTGTGAAGGCAAGAGTGCCTACAGATCTGTAACTT 1110
QY 1702 GAGCCCTGCCCCAGCTCAGCCCTCCGGAAGAGCTTCCGGGAGGAGCAGTGTGAGGCTTTC 1761
DB 1111 GAGGACTGTCCAGC---AATAATGGAAGAACCTTTAGAGAGGAACAATGTGAAGCACAC 1167
QY 1762 AACGGCTACAACCAAGCAGCAACCGGCTCACTCTGCGCGTGGCATGGGTGCCCAAGTAC 1821
DB 1168 AACGAGTTTTCAAAAGCTTCTCTTGGGAGTGGGCTGCGGTTGGAATGGATTTCCCAAGTAC 1227
QY 1822 TCCGGCTGTCTCCCGGAGCAAGTCAAGCTCATCTCCGAGCAATGGCACTGGCTAC 1881
DB 1228 GCTGGCTCTCAACAAAGCAGGTGCAAGTCACTCTCCCAAGCCAAAGGATTTGGCTAC 1287
QY 1882 TTCTATGTCTGGCCACCAAGTGTGTGACCGCAGCTGTGCTCTCTGATCTCCACTTCC 1941
DB 1288 TTCTTCGTTTTCAGCCCAAGGTTGTTGATGTTACTTCCATGTAGCCAGATTTCCACTCT 1347
QY 1942 GTCTGTGCCAGGCAAGTGCATCAAGCTGGCTGTGTGAGGAACTGGGCTCCCAAGAG 2001
DB 1348 GTCTGTGTGCAAGGACAGTGTGTAAGCTGGTGTGTGATCGCATCATAGCTTCCAAAAG 1407
QY 2002 AGATTGCAACAGTGTGGGTGTGTGGGGAGACAAATAAGAGCTGCAAGAAGTGTGACTGGA 2061
DB 1408 AAGTTGATAATGTGGTGTGTTGGCGGGGAATGGATCTACTTGTAAATAATATACAGA 1467
QY 2062 CTCTTCAACAGCCCATGATGGCTTACAATTTGTTGGTGGCCATCCCGCAGGCGCTCA 2121
DB 1468 TCAGTTACTAGTGCAAAACCTGGATATCATGATATCGTCACAATTTCCAACTGGAGCCACC 1527
QY 2122 AGCATCGACATCCCGCAGCGGTTTACAAGGCTGTATGGGGATGAGCAACTACCTGCT 2181
DB 1528 AACATCGAAGTGAACAGCGGAACCCAGAGGGGATCCAGGAACAATGGCAGCTTCTTGCC 1587
QY 2182 CTGAAGAACAGCAAGGCAAGTACCTGCTCAACGGGCAATTTCTGTTGTGCGCGGTGAG 2241
DB 1588 ATCAAGCTGTGTATGGCAGATATATCTTATGTGTACTACACTTGTCCACTTAGAG 1647
QY 2242 CGGGACCTGTGTGAAGGCGAGTCTGCTGGGTATACGCGGACGGGCGACAGCGGTGAG 2301
DB 1648 CAAGACATTATGTACAAAGTGTGTCTGTAGGTACAGCGGCTCTCTGCGGCAATGGAA 1707

Qy	1051	GATGGGTTCCATCAGCTTCAACCTGCCACGAGCTGGCCAGCTGTTCACATGCC	1110
Db	451	GATGGTTTAAAGTGCCTTCACACAGCCCTGAATATAGCCAGCTGTTTAACTGCCA	510
Qy	1111	CATGACAATGTGAAGCTCTGTGAGGAGTGTGTTGGGAAGCTCCAGCCAAACCATGATG	1170
Db	511	CATGATGATCGCAAGCAGTGTGCCAGCCTTAATGTGTGTGAACAGGATTCACATGATG	570

Qy	1171	TC	CCGACG	CCCTCAT	CCAGAGAT	CGACCG	CGCAAC	CCCTGGT	CAGCTG	CGAGT	CGTGC	CAATC	1233
Db	571	CG	GTCAAT	CGTTT	CCCAAC	CTGGAC	CACAGC	CCCTT	GGTCT	CTCCT	TCGAGT	GCCTAC	630
Qy	1231	AT	CACGAC	TTCT	CGACAC	CGGGC	ACGGT	GACTG	CTCT	CTCGG	ACCAAC	CCGCAAGCC	1290
Db	631	ATT	ACAT	CATT	TTCT	TGGAT	AATG	CTAT	GGGAAT	GT	TTGT	TGGC	690
Qy	1291	AT	CTCC	CTG	CCCG	CAGAT	CT	CGCG	GGCG	CCAG	CTAC	ACCC	1350
Db	691	AT	CAGCT	CC	CAGCG	AT	CT	CT	CG	AC	CT	TT	750
Qy	1351	G	CTTT	TGG	CT	TGG	CT	TCC	AAG	CCCT	GT	CT	1407
Db	751	AC	ATT	TGG	GAG	GACT	CC	AAAC	CAC	TGCC	CGAT	GC	810
Qy	1408	TG	CAC	CGG	GAAG	CC	AAG	GC	ACAG	AT	GT	GT	1467
Db	811	TG	TAC	CGGC	AC	CT	CT	GGT	GGGT	GT	GT	GT	870
Qy	1468	GG	CAC	AG	CT	GT	GG	CAG	GGC	AA	GT	CT	1527
Db	871	GG	CAC	AG	CT	GT	GG	GAAG	GGAA	AT	GGT	GT	930
Qy	1528	CT	CA	CA	AG	CA	AG	-----	GT	GAT	GGT	TT	1581
Db	931	AG	GA	AG	CAT	TTT	T	GAT	AC	GC	CT	TT	990
Qy	1582	TG	CT	CG	GC	AC	AT	GT	GGG	GGC	GT	GC	1641
Db	991	TG	TT	CG	GA	AC	GT	GT	GGG	GT	CA	CT	1050

QY	1642	CTTGCCACCGGGGCAAGTACTCGCAGGAGTGAGGTGGAATACCGATCTGCAACCTG	1701
Db	1051	CCAAAGAATGGAGGGAAGTACTGTCAAGGCAACAGGTGCGCTACAGATCCGTGAACCTT	1110
QY	1702	GAGCCCTGCCCAAGTCAGCCTCCGGAAGAAGAGCTTCCGGGAGGAGCAGTGTGAGGCTTTC	1761
Db	1111	GAGGACTGTCCAGAC---AATAATGGAAAAACCTTTAGAGAGGAACAATGTGAAGACAC	1167
QY	1762	AACGGCTACAAACACAGACACCAACACGGCTCACTCTCGCGGTGGCATGGGTGCCCAAGTAC	1821
Db	1168	AACGAGTTTTCAAAGCTTCCTTTGGGAGTGGGCTGGGTGGAATGGATTTCCAAGTAC	1227
QY	1822	TCCGGGGTGTCTCCCGGGGACAAGTGCAAGTCTATCTGCGAGGCCAATGGCACTGGCTAC	1881
Db	1228	GCTGTGGGCTCTACCAAGAGACAGTGTCAAGTCTATCTGCCAAGCCAAGGCATTTGGCTAC	1287
QY	1882	TTCTATGTCTGTGGCACCCAAAGTGTGTGACGCGCAGGTGTGCTCTCTCTGACTCCACCTCC	1941
Db	1288	TTCTTCGTTTTCGACGCCAAGGTGTGATGGTACTCATGTAGCCCAGATTTCCACCTCT	1347
QY	1942	GTCTGTGTCCAAAGCAAGTGATCAAGGCTCGGCTGTGATGGAACTTGGGCTCCAAAGAAG	2001
Db	1348	GTCTGTGTGCAAGGACAGTGTGTAAAGCTGTTTGTGATCGCATCATAGACTCCAAAAAG	1407
QY	2002	AGATTCGACAAAGTGTGGGTGTGTGGGGGAGACAAATAGAGCTCCAAAGAGTGACTGGA	2061
Db	1408	AAGTTTGATAATGTGGTGTTCGGGGGGAATGGATCTACTTGTAAAAAAATATCAGGA	1467
QY	2062	CTCTTCACCAAGCCCATGCATGGCTTACAAATTTCTGTGTGGCCATCCCGCGAGCGGCTCA	2121
Db	1468	TCAGTTTACTAGTGTCAAAACCTGGATATCATGATATCGTCACAATTCACATCGAGCCACC	1527

QY 2122 AGCATGCACATCCGCCAGCGGTTTACAAAGGCGTGATCGGGGATGACAACTACCTGGCT 2181
 DB 1528 AACATCGAAGTGAACAGGGAACAGGAGGATCCAGGAACAATGCGAGCTTCTTGCC 1587
 QY 2182 CTGAAGAACAGCCAGCAAGTACCTGCTCAACGGGCAATTCGTGGTGTGCGGGTGGAG 2241
 DB 1588 ATCAAAGCTGCTGATGTCACATATATCTTAATGGTGACTACACTTTGTGCCACCTTAGAG 1647
 QY 2242 CGGACCTGTGTGAGGAGCGTCTGCTCGGTGACAGCGGCGGACAGCGGTGGAG 2301
 DB 1648 CAAGACATATGACAAAGTGTGCTTGAGGTACAGCGGCTCTCTGCGGCAATGGAA 1707
 QY 2302 AGCCTGAGGCTTCCCGGCCCATCTCGGAGCGCTGACCGTGGAGTCTCTCCGTTGGG 2361
 DB 1708 AGAATTCGAGCTTTAGCCCTCTCAAGAGCCCTTGACCATCCAGGTCTCTTACTGTGGG 1767
 QY 2362 AAGATGACACCGCCCGGCTCGGTACTCTTCT 2395
 DB 1768 AATGCCCTTGACCTAAATAAATAACACTACT 1801

RESULT 12
 US-09-445-023A-2
 ; Sequence 2, Application US/09445023A
 ; Patent No. US20020119167A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hirose, Kunitaka
 ; APPLICANT: Inoguchi, Ei-ji
 ; APPLICANT: Hakozaaki, Michinori
 ; APPLICANT: Ishioaka, Keiko
 ; APPLICANT: Ishida, Yukako
 ; APPLICANT: Matsushima, Kouji
 ; APPLICANT: Kuno, Kouji
 ; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
 ; FILE REFERENCE: 057092
 ; CURRENT APPLICATION NUMBER: US/09/445,023A
 ; CURRENT FILING DATE: 1999-12-03
 ; PRIOR APPLICATION NUMBER: JP 9-160422
 ; PRIOR FILING DATE: 1997-06-03
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 2184
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: exon
 ; LOCATION: (1)..(2184)
 ; US-09-445-023A-2

Query Match 22.0%; Score 628.8; DB 10; Length 2184;
 Best Local Similarity 60.9%; Pred. No. 2.4e-158;
 Matches 1080; Conservative 0; Mismatches 682; Indels 12; Gaps 3;

QY 631 AAGCGTTTCGTCTATCCCGCGGTAGCTGGAGACGCTGGTGGTGGCGGAGAGTCAATG 690
 DB 31 AAGCGTTTCGTCTATCCCGCGGTAGCTGGAGACGCTGGTGGTGGCGGAGAGTCAATG 90
 QY 691 GTCAGTTCCACGGCGGACCTGGAACATTTATCTGCTGAGCTGCTGGCAACGGCGGCG 750
 DB 91 GCAGATTCCACGGCAGTGGTCTTAAGCATTTACCTTCTCAGCTGTTTTCGGTGGCAGCC 150
 QY 751 CGACTTACCAGCTCCAGCATCTCAACCCCATCAACATCGTTGTGGTCAAGTCTG 810
 DB 151 AGATTGTACAACACCCAGCATCTGTAATTCAGTTAGCTGGTGGTGGTGAAGATCTG 210
 QY 811 CTCTTAGAGATGCTGACTCCGGGCCCAAGTCAACCGCAATCGGCGCTGACGCTCGC 870
 DB 211 GTCATCCAGATGAACAGAGGGCGGAGTACCTCCATGCTGCCCTCACTCTCGG 270
 QY 871 AACTTCTGCTGCGCAGCAAGCTGAACAAAGTGAGTGACAGACACCCCGAGTACTGG 930

DB 271 AACTTTTGCACATGCGAAGAGCAGACAAACCCACCTGACCGGGATGCAAGCACTAT 330
 QY 931 GACACTGCCATCTCTTACCAGGAGGACCTGTGTGAGGACCAACCACTGTGACACCTGT 990
 DB 331 GACACAGCAATTCCTTTACCAGAGAGCTGTGTGGTCCAGACATGAGTACTCTT 390
 QY 991 GGCATGGCTGATGTGGGTACCATGTGTGACCCCAAGAGAGTGTCTGTCTATTTGAGGAC 1050
 DB 391 GGGATGGCTGATGTGGAACTGTGTGATCCGAGCAGAGTGTCTGCTCATAGAAGAT 450
 QY 1051 GATGGCTTCCATGACCTTCCACACTGCCACAGCTGGGCCAGCTGTCTCAACATGCCC 1110
 DB 451 GATGGTTTAAAGCTGCGCTTCCACAGCCCATGAATTAAGCCACGCTGTTTAAACATGCCA 510
 QY 1111 CATGACAATGTGAAAGTCTGTGAGGAGGTGTGGGAAAGCTCCGAGCCAAACACATGATG 1170
 DB 511 CATGATGATGCAAGCAGTGTGCCAGCCTTAATGCTGTGAACAGGATTCACCATGATG 570
 QY 1171 TCCCGACCTCATCCAGATCGACCGTGCACACCCCTGGTCAGCTGCTGCTGCCATC 1230
 DB 571 GCGTCAATGCTTTTCCAACTGAGCCACAGCCAGCCTTGGTCTCTCTGAGTGCCTACATG 630
 QY 1231 ATCACCAGCTTCTTGACAGCGGCGAGCTGACTGCTCTGACCAACCCAGCAAGGCC 1290
 DB 631 ATTACATCAATTTCTGGATTAATGCTATGGGAATGTTTGTGAGCAAGCCATCAGATCCC 690
 QY 1291 ATCTCCCTCCCGAGGATCTCCGGGGCGGAGCTACACCTGAGCCAGAGTGCAGTGTG 1350
 DB 691 ATACAGCTCCAGGGGATCTCCCTGCGACCTTGTACGATGCCAACCCGCGAGTGCAGTTT 750
 QY 1351 GCTTTTGGCGTGGCTCCCAAGCCCTGTCTTACATGCAGTAC--TGACCAAGCTGTGG 1407
 DB 751 ACATTTGGGAGAGCTTCCAAACACTGCCCCGATGAGCCAGCAGATGTAGCACTTTGTGG 810
 QY 1408 TGCAACCGAAGGCCAAGGAGCAGATGTGTGCCAGACCCGCACTTCCCTTGGGCCGAT 1467
 DB 811 TGTACCGGACCTCTGTGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 870
 QY 1468 GGCACAGCTGTGGGAGGCGGAGCTGTGCTCAAGAGGCGCTGTGTGAGAGACACACAC 1527
 DB 871 GGCACAGCTGTGGAGAGGGAATGTGTATCAACGGCAAGTGTGTGTGTGTGTGTGTGTGT 930
 QY 1528 CTCACAACGACACAGG-----GTGGATGTTCTCTGGGCGCAATGGATCCCTTATGGCCCC 1581
 DB 931 AGGAAGCTTTTGTATAGGCTTTTTCATGGAAGCTGGGAGCCTATGGGACCTTGGGAGAC 990
 QY 1582 TGCTCGCGCACATGTGTGGTGGGCGGTGACAGCTGGCGAGGAGCAGTGCACCAACCCAGC 1641
 DB 991 TGTTCGAGAAGCTGCGGTGGAGAGTCCAGTACACGATGAGGGAATGTGACAAACCCAGTC 1050
 QY 1642 CTGCGCAACGGGCGCAAGTACTGCGAGGAGTGGGGTGAATACCGATCTCTGCAACCTG 1701
 DB 1051 CCAAGAATGGAGGAGTACTGTGAAGCAACAGCTGCTACAGATCTCTGTAACTTT 1110
 QY 1702 GAGCCTGCCCCAGCTCAGCCTCCGGAAAGAGCTTCCGGAGGAGCAGTGTGAGGCTTTC 1761
 DB 1111 GAGGACTGTCCAGAC---AATTAATGGAAGAACTTTAGAGAGCAACATGTGAGCAGAC 1167
 QY 1762 AAGCGGTACAACACAGCAGCAACCCGCTCACTCTGCGCGTGGCATGGGTGCCCAAGTAC 1821
 DB 1168 AACGAGTTTTCAAAAGCTTCTTTTGGGAGTGGGCGCTGCGGTGGAATGGAATCCCAAGTAC 1227
 QY 1822 TCCGGGCTGTCTCCCGGACAGATGCAAGCTCATCTGCCGAGCCAACTGGCATCTGCTAC 1881
 DB 1228 GCTGGGCTGTCAACAAAGAGCAGGTGCAAGCTCATCTGCCAAGCCAAAGCAATGGCTAC 1287
 QY 1882 TTCTATGTGTGCGCAGCCCAAGTGTGTGGAGCGGAGCAGCTGTGCTCTCTGCTGACTCCACTCC 1941
 DB 1288 TTCTCGTTTTCGAGCCCAAGGT 1347
 QY 1942 GTCTGTGTCAAGGCAAGTGCATCAAGGCTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2001
 DB 1348 GTCTGTGTGCAAGGACAGT 1407

QY	2002	ATGATTGCACAGTGTGGGTGTCTGTGGGGAGACAATAAGAGCTGCAAGAAGGTGACTGGA	2061
Db	1408	AAGTTTGATAAATGTGTGTCTTCGCGGGGAAAATGGATCTACTTGTAAAAAATATCAGGA	1467
QY	2062	CTCTFTACCAGCCCATGATGCTACAAATTTCTGTGTGGCCATCCCGCAGGCGCCTCA	2121
Db	1468	TCAGTTACTAGTGCANAACTGGATATCAATATCGTCAAAATTCCAAATGGAGCACCC	1527
QY	2122	AGCATCGACATCGCCGAGCGCGTGTACAAAGGGCTGATCGGGGATGACAACTACCTGGCT	2181
Db	1528	AACATCGAAGTGAACAGCGGAACACAGAGGGGATCCAGGAACAATGGCAGCTTCTTTGCC	1587
QY	2182	CTGAAGAACAACGCCAAGSCAAGTACCTGCTCAACGGGCATTTCTGTGTGTCTGGCGGTGGAG	2241
Db	1588	ATCAAAGCTCTGATGGCACATATATCTTAATGTGACTACACTTTGTCACACTTAGAG	1647
QY	2242	CGGGACCTGCTGTGAAGGCGAGTCTGCTGCGGTACAGCGGCAGGCGACAGCGGTGGAG	2301
Db	1648	CAAGACATTAATGTACAAAGGTGTGTCTTTGAGGTACACGGCTCCTCTCGGCATTCGAA	1707
QY	2302	AGCCTGCAGCCTTCGCGGCCCATCTTGGAGCGCGCTGACCGTGGAGGTCTCTCCGTGGGG	2361
Db	1708	AGAAITCGCAGCTTTAGCCCTCTCAAGAGOCCTTGACCATTCCAGTTCTTACTGTGGGC	1767
QY	2362	AAGATGACACCGCCCGCGGTTCGCGTACTCCTTCT	2395
Db	1768	AATGCCCTCGACCTAAATAATACACCTACT	1801
 RESULT 13 US-10-105-929-1 ; Sequence 1, Application US/10105929 ; Patent No. US20020137142A1 ; GENERAL INFORMATION: ; APPLICANT: Holtzman, Douglas A. ; APPLICANT: Goodearl, Andrew D.J. ; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83 ; FILE REFERENCE: 09404/041001 ; CURRENT APPLICATION NUMBER: 2002-03-25 ; CURRENT FILING DATE: 2002-03-25 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/130,491 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108 ; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961 ; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06 ; NUMBER OF SEQ ID NOS: 16 ; SOFTWARE: FastSeq for Windows Version 3.0 ; SEQ ID NO 1 ; LENGTH: 4676 ; TYPE: DNA ; ORGANISM: Homo sapiens ; FEATURE: ; NAME/KEY: CDS ; LOCATION: (460)...(3360) US-10-105-929-1			

Query Match 22.0%; Score 628.2; DB 12; Length 4676;
Best Local Similarity 57.1%; Pred. No. 3.9e-158;
Matches 1305; Conservative 0; Mismatches 913; Indels 69; Gaps 16;

[illegible]

QY	352	ACGCTGTGCGGGGGCTCCCGCGAGCCTTTGGTACCGAGGCCGAGTATGTCATTAGC	411
Db	877	AGCCTCTCGGAGGGCTGCGCGGCGCTTCTACCTGCTGGGGAGCGTATTTCATCCAG	936
QY	412	CGCTGCGCAATGCTAGCGGC -CGGCGGCGAGCGACACAGCCAGGGCGCACACCTTCT	470
Db	937	CGCTGCGCGCGCAGCGAGCGCTCGCCACCGCGCGCCACGAGGAGGAAGCCGCGGCA	996
QY	471	CGAGCGCGGGGTGTTCCGGCGGGCGCTTCCGGAGACCCACCTCTCGCTCGGGGTGGC	530
Db	997	CCACTACAGTTCACCTCTCGCGGGAATCGCGAGGCGACGTAGGCGGCGACGTGCGGG	1056
QY	531	CTCGGGCTGGAAACCCGCCATCTACGGGCCCTTGAGCCCTTACAAGCCGCGCGGGCGG	590
Db	1057	GTCTGGACGACGAGCCCCCGCGACTGGGAAACGGAGACCGCAAGACGAGACGAAGG	1116
QY	591	CTTCGGGAGAGTCTGTAAGCCGGGCGAGGTCGTGGCGC -----	627
Db	1117	ACTGAGGCGAGGACGAAGGCGCTCAGTGTGTCGCCGACGAGCCCGCACTGCAAGCGTA	1176
QY	628	-----GCCACGCTTTCGTGCTATATCCCGGGTAC	657
Db	1177	GGACAGCCCCAGGAACCTGGAAGCATAGAAGAAAGCGATTGTCTCCAGTCAACGCTAT	1236
QY	658	GTGAGACGCTGTGTGTCGCGGACGAGTCAATGGTCAAGTTCACGGCGCGGACCTGGAA	717
Db	1237	GTGGAACCATGCTTGTGGCAGACCAGTCGATGGCGAGATTCACGGCAGTGGTCTAAG	1296
QY	718	CATTATCTGTGAGCTGTGTGGCAAGGGGGCGACTCTACCGCATCCACGACATCCTC	777
Db	1297	CATTACCTCTCAGTGTGTTTCGGTGGCAGCAGATTGTACAAACACCCAGCATTCGT	1356
QY	778	AACCCCATCAACATCGTTGGTCAAGTCTCTTTAGAGATCTGTACTCCCGGCC	837
Db	1357	AATTGAGTTAGCCTGTGTGTGGTGAAGATCTTGGTCATCCAGCATGAACAGAAGGGCGG	1416
QY	838	AAGTCAACCGCAATCGGCCCTGACGCTGCGCACTCTGTGCTCGGCAGAAAGCTG	897
Db	1417	GAGTGAACCTCAATGTCGCCCTCACTCTGCGGAACCTTTTGCACTGGCAGAAGCAGC	1476
QY	898	AACAAAGTGATGACAGCACCCGAGTACTGGGACACTGCCATCCTCTTCAACAGGCAG	957
Db	1477	AACCCACGAGTGACGGGATGCGAGGCACTATGACACAGCAATTTCTTTCAACAGACAG	1536
QY	958	GACCTGTGTGAGCCACCACTGTGACACCCCTGGGCATGGCTGATGTGGGTACCATGTG	1017
Db	1537	GACTGTGTGGTCCCAGACATGTGATACTCTTGGATGGCTGATGTTGGAACGTGTGT	1596
QY	1018	GACCCCAAGAGCTGCTGTCTATTGAGGAGGATGGCTTCCATGACGCTTCAACACT	1077
Db	1597	GATCCGAGCAGAGCTGCTCCGTCATAGAAGATGATGTTTACAAGCTGCCTTCAACACA	1656
QY	1078	GCCACGAGCTGGGCCACGCTGTTCACATGCCCATGACAAATGTGAAGTCTGTGAGGAG	1137
Db	1657	GCCCATGAATTAGSCCAGCTGTTACATGCCACATGATGATGCAACAGCTGTGCCAGC	1716
QY	1138	GTGTTTGGGAGCTCGAGCAACCACTATGTCCCCGAGCCTCATCCAGATCGACCGT	1197
Db	1717	CTTAATGGTGTGAACAGGATTCACATGATGGGCTCAATGCTTTCCAACTGGACAC	1776
QY	1198	GCCAAACCCCTGTGACCTGCACTGCCATCATCCGACTTCTTGGACAGCGGCAC	1257
Db	1777	AGCAGCCTTGGTCTCGTTGAGTGCTACATGATTATCATTTCTGGATTAATGGTCTAT	1836
QY	1258	GGTGACTGCTCTCTGGACCAACCCAGCAAGCCCATCTCCCTGCCCGAGGATCTGCCGGC	1317
Db	1837	GGGGAATGTTTGTATGGACAAGCCTCAGAAATCCCATACAGCTCCCAAGCGATCTCCCTG	1896
QY	1318	GCCAGCTACACCTGTAGCCAGCAGTGCAGCTGCGCTTTTGGCTGGGCTCCAGCCCTGT	1377
Db	1897	ACCTCTGAGTATGCCAAACCGGACGTCCAGTTTACATTTATTTGGGAGGACTCCAACACTGC	1956


```
Db 1037 CTGCAGAACACATCTCTCAGGTGATGTCATATGCGACCCGATCTACAGCACCCGAC 1096
QY 772 ATCTCAACCCATCAACATCTGTTGGTCAAGGTCTGCTTCTTAGAGATCGTGACTCC 831
Db 1097 ATCAGGAACCTCCGTCACCTTGTGGTGGTGAAGTCTAATAGTGGAAGAAAGATGG 1156
QY 832 GGCCCAAGGTCAACCGCAATCGGCCCTGAGCTCGCAACTTCTGTGCTCGGAGAG 891
Db 1157 GGCCCGAAGTGTCCGACACACGGGGGCTCACACTCGGCAACTTTCGACGTGGCAACGG 1216
QY 892 AAGCTGAACAAAGTAGTGACAAAGACACCCGAGTACTGGGACACTGCCATCTTTCACC 951
Db 1217 CGTTTCAACAAGCCCAAGTACCGCCACCCGAGCACTATGACACTGCCATCTTGTTCACC 1276
QY 952 AGCAGGACTGTGTGG---AGCCACCCTGTGACACCTGGCATGGCATGCTGTGGGT 1008
Db 1277 AGACAAACTTGTGGGAGGAGAGAGTGTGACACCTGGGATGGGACAGCTTGGC 1336
QY 1009 ACCATGTGTGACCCCAAGAAAGTCTCTGTCAATTGAGGACGATGGGCTTCCATCAGCC 1068
Db 1337 ACCATGTGTGACCCGACAAAGAGTCTCTCAGTGATCAAGGATGAGGACTGAGCAGCC 1396
QY 1069 TTCACCACTGCCACAGCTGGCCACAGTGTTCACATGCCCATGACAAATGTGAAGTC 1128
Db 1397 TACACCTGTGGCCATGAGTAGGGACGTTCTCAGCATGCCCATGATTTCTAAGGCC 1456
QY 1129 TGTGAGGAGGTGTTTGGGAAGTCCGAGCCAAACACATGATGTCCCGACCTCATCCAG 1188
Db 1457 TGTGTGAGATTGTTGGGCCATGGGCAAGTACACATGATGGGCCATCTTTCATCCAC 1516
QY 1189 ATCGACCTGGCAACCCCTGTGAGCTGAGCTGCTGCCATCATCAGGACTTCTGGAC 1248
Db 1517 GTGAACAAGACGCTGCGCTGTCTCCTGCAGTGTCTACCTCACAGAGTCTCTGAT 1576
QY 1249 AGCGGCAGGTGACTGCTCTCGACCAACCCAGCAAGCCCATCTCCTGCCGAGAT 1308
Db 1577 GATGTCAGGAGATTGTTCTGATGCCCCACCTCGGTCTGCCCTCCCCACAGCC 1636
QY 1309 CTGCGGGGCCAG-----CTACACCTTGAGCCAGAGTGCAGCTGGCTTTTGGCGTG 1362
Db 1637 CTCCGGGCCACAGCACCCCTCTACGAGCTGGACAGCAGTGCACAGCATCTTGGGCT 1696
QY 1363 GGCTCCAGCCCTGCTTACAG-TGCACTGACCAAGCTGTGGTGGACCGGGAAGCC 1421
Db 1697 GATTTCCGACACTGCCCAACACCTCTGTGGAGGACATCTGTGACAGCTGTGCCCGT 1756
QY 1422 CAAGGACAGATGTTGTGCC-----AGCCCGCACTTCCCTGGGCGCAT 1467
Db 1757 CATCGGATAGTGTAGGCCCATTTGCCACAAAGATGTGAGCTGCTCTGGGCTGAT 1816
QY 1468 GGCACAGCTGTGGGAGGCAAGCTCTGCCTCAAAGGGGCTCGGTGGAGAGACACAAAC 1527
Db 1817 GGTACACCTGTGGGCTGGGACCTGTGCCTGGATGTAGTGTGTACTCAAGGAGAT 1876
QY 1528 CTCACAAGCACA-----GGGTGATGTTCTCGGCCCAATGGGATCCCTATGGCCCC 1581
Db 1877 GTGGAGATCCCAAGCTGTGTAGTAGAGACTGGGTTCCTGGAGACCTGGGGGACAA 1936
QY 1582 TCCTCGGCACATGTGTGGGGGCTGCGCTGGCCAGGAGGAGTGCACCAACCCACC 1641
Db 1937 TGTTCGACCTGTGGTGGAGGATACAAATCTCGAACCTGATGTGATATCCCAATG 1996
QY 1642 CTGCGCAAGGGGCAAGTACTGCGAGGAGTGGGTGAATACCGATCCTGCAACCTG 1701
Db 1997 CCTCAGATGGAGGAAGATTTCCTGCTGAGTGAAGTCAAGTACCAATCATGCAACACA 2056
QY 1702 GAGCCCTGCCAGCTGACCTCGGAAAGAGCTTCCGGGAGGAGCAGTGTGAGGCTTC 1761
Db 2057 GAGGAATGTC-----CACCACCAAGGAAAAAGCTTCCGGGAGCAGCAGTGTGAGAAATAT 2110
QY 1762 AACGGCTACAAACACAGCACCAACCGGCTCACTCTCGCGGTGGCATGGTGGCCCAAGTAC 1821
```

```
Db 2111 AATGCCTACARCCA---CACTGACCTGGATGGGAATTTCTGCAGTGGTCCCAAGTAT 2167
QY 1822 TCCGGCTGTCTCCCGGGGACAAAGTGCAAAGTCTATCTGCCGAGCAATGCTACTGGGTAC 1881
Db 2168 TCAGGAGTGTCCCGCCGAGACCGATGCAAGCTGTTTTGSCAGAGCCGTTGGAGG-NGTGA 2227
QY 1882 TTCTATGTCTGGCACCCCAAGGTGGTGGACGGCAGCTGTCTCTGACTCCACCTCC 1941
Db 2228 TTCAAAGTGTTTGAAGCTTAAGGTGATCGATGGCACCTGTGTGTGGACGGGATACTCTGTCC 2287
QY 1942 GTCTGTGTCCAAAGCAAGTGCATCAAGGCTGGCTGTGATGGGAACCTGGGCTCCCAAGAAG 2001
Db 2288 ATCTGGTCCGGGGCAATGTGTTAAGGCTGCTGTGACCATGTGTGAATCACTTAAG 2347
QY 2002 AGATTGCAAGTGTGGGGTGTGTGGGGGAGACAATTAAGAGCTCCAAAGAAGTGAAGTGA 2061
Db 2348 AAGCTGGAACAATGTGGGGTGTGTGGGGGCAAAAGGCATGCTCTAGGAAGATCTCCGGT 2407
QY 2062 CTCCTTACCAAGCCCATGATGGCTACAATTTCTGTGTGGCCATCCCGCAGGCGCTCA 2121
Db 2408 TCTTTCACCCCTTTCAGTTATGGCTACAATGACATTTGTACCATCCAGCTGGTGCACA 2467
QY 2122 AGCATCGACATCCCGCAGCGGTTTACAAAGGGCTGATCGGGGATGACAACATACTGGCT 2181
Db 2468 AACATTGATGTGAACAGCGGAGTCAACCAGGGGTCAAGAACAGCGCAGCTACTGGCG 2527
QY 2182 CTGAAGACAGCCAAAGCAAGTACTGCTCAACGGGATTTCTGTGTGGTGGTGGGAG 2241
Db 2528 CTGAAGACAGCCAAATGGGCAATGCTCTCAATGGTAACTTGGCCATCTCTGCCATAGAG 2587
QY 2242 CGGACCTGTGTGTAAGGGCAGTCTGTGCGGTACAGCGCAGCGGACAGCGGTGGAG 2301
Db 2588 CAAGACATCTGTGTGAAGGGACCATCTTGAAGTACAGTGGCTCCATGGCTACCTGGAG 2647
QY 2302 AGCTGAGGCTTCCCGGCCCATCTTGGAGCCGCTGACCGTGGAGGTCCTCTCGTG--- 2358
Db 2648 CGGTGACAGAGCTTCCAGGGCCTGCTGAGCCTCTTACAGTACAGCTCTGACTGTCT 2707
QY 2359 GGAAGATGACACCCCGGGCTGCTACTCTCTCTATCTGCCCCAAGA 2408
Db 2708 GGTGAGGTCTTCCCTCCAAAGTCAGATATACCTTTTGTGCCCAATGA 2757
```

RESULT 15

```
US-10-174-590-351
; Sequence 351, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 351
; LENGTH: 4407
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-174-590-351
```

Query Match

20.4%; Score 583.2; DB 9; Length 4407;

Best Local Similarity 60.4%; Pred. No. 4, 1e-146;		Matches 1093; Conservative 0; Mismatches 693; Indels 24; Gaps 57;	
QY	609	CCGGCGAGGTGTGGCGCGCCNAGGGTTTGGTGTATCTCCCGCGGTAGCTGGAGACGCT	668
Db	1027	CAGCCCGACACCCGAGAGCCGCTTCTGCTTACGATAGTAGATTTGTGGAGACACT	1086
QY	669	GGTGGTGGGAGGAGTCAATGGTCAAGTTCACGCGCGGACCTGGAACATTTATCTGCT	728
Db	1087	GGTGGTGGGAGATGACAAGATGGCCGATTTCCAGGTGGGGCTTAAGCGCTACTGCT	1146
QY	729	GAGCGTGTGGCAACGCGCGGACTCTACCGCCATCCAGCATCTTCAACCCCATCAA	788
Db	1147	AACAGTGTGGCAGCAGCAGCCCTTCAAGCACCCCAAGCATCCGCAATCTGTAC	1206
QY	789	CATCGTGTGGTCAAGGTGCTGCTTTAGAGATCGTACTCCGGGCCCAAGGTCAACGG	848
Db	1207	CTTGGTGGTCACTCGGCTAGTGTCTGGGGTACGCGAGGAGGGCCCAAGTGGGGC	1266
QY	849	CAATGGGGCCCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAAGCTGAACAAAGTGA	908
Db	1267	CAGTGTGCCCGACCCCTGGCAGCTTCTGTGCTGGCAGCGGGGCTCAACACCCCTGA	1326
QY	909	TGACAGACCCCGAGTACTGGGACACTGCGATCTCTTCAACAGCAGGACCTGTGTGG	968
Db	1327	GGACTCGGGCCCTGACCACTTTGACACAGCCATTCTGTTTACCCGTCAGGACCTGTGTGG	1386
QY	969	AGCCACACCTGTGACACCTCGGCGTGGTGTGATGGGTACCATGTGTGACCCCAAG	1028
Db	1387	AGTCTCACCTTGGACACGCTGGGTATGGCTGATGTGGGACCGCTGTGTGACCCGCTG	1446
QY	1029	AAGTGTCTGTCTATTGAGGACGATGGGTTCCATCAGCCTTCAACACTGCCCAAGCT	1088
Db	1447	GAGCTGTGCCATTGTGGAGGATGATGGGTCCAGTCCAGCTTCACTGTCTCATGAAT	1506
QY	1089	GGGCGAGCTGTTCAACATGCCCCATGACATGTGAAGTCTGTGAGGAGGTGTTGGGA	1148
Db	1507	GGGTCTATGTTTCAACATGTCTTCAACATGTCTTCAACATGTCTTGAATGGGGC	1566
QY	1149	GCTCCGAGCCAA---CCACATGATGTCCCGACCCCTCATCCAGATCGACCGTGCCCAACCC	1205
Db	1567	TTTGAGCACCTCTCGCCATGTCATGGCCCTGTGATGGCTCATGTGATCCTGAGGAGCC	1626
QY	1206	CTGCTAGCCTCGAGTGTGCCATCATACCGACTTCTTGGAGAGCGGGACGGTGAATG	1265
Db	1627	CTGTGCCCCCTGACGTGCCGCTTCACTAGTACTTCTTGGACAAATGGCTATGGGCACTG	1686
QY	1266	CCTCTGGACCAACCCAGACGCCATCTCCCTGCGGAGGATCTCGCGGGCGCCAGCTA	1325
Db	1687	TCTCTTAGAACAAACAGAGGCTCCATGTGATCTGCTGTGACTTTCCCTGGCAAGACTA	1746
QY	1326	CACCTGAGCGCAGCTGCGAGCTGGCTTTTGGCTGGGCTCCAAGCCCTGTCTTACAT	1385
Db	1747	TGATGCTGACCGCCAGTGCCAGTGACCTTCCGGCCGAGCTCACGCCATTGTCCACAGCT	1806
QY	1386	---CGAGTACTGACCAAGCTGTGGTCCACCGGAGGCCAAGGACAGATGTGTGCCA	1442
Db	1807	GCCCGCCCTGTGCTGCTCCCTTGTGCTCTGGCCACCTCAATGGCCATGCCATGTGCCA	1866
QY	1443	GACCGGCACCTTCCCTGGCGCGATGACGCTGTGGCGAGGCAAGCTCTGCGCTCAA	1502
Db	1867	GACCAACACATCGCCCTTGGCCGATGGACACACCTTGGGGCCCGCACAGCCCTGCATGGG	1926
QY	1503	AGGGGCTCGGTGGAGAGACACAACTTCAACAGCACAGGGT-----GGATGGTTCCTG	1556
Db	1927	TGCTGCTGCTCCACATGGACAGCTCCAGGACTTCAATATTTCCACAGGCTGTGTGCTG	1986
QY	1557	GGCCAAATGGATCCCTATGGCCCTCTGCGCACATGTGTGGGGCGGTGCGAGCTGGC	1616
Db	1987	GGGTCTTGGGGACCATGGGGTACTCTCTGGACCTGTGGGGTGGTGTCCAGTTCTC	2046
QY	1617	CAGGAGGAGTGCACCAACCCACCCCTGCGCAACGGGGGCAAGTACTTGCAGGAGTGA	1676

...

Db	2047	CTCCCGAGACTGCACGAGGCGCTGTCCCGGGAATGGTGGCAAGTACTGTAGGCGCCGCG	2106
QY	1677	GGTGAATACGATCTCTGCAACCTTGGAGCCCTGCCAGCTCAGCCTCCGGAATAGACTT	1736
Db	2107	TACCGCTTCCGCTCTCTGCAACACTGAGGACTGCCCAA---CTGGCTCAGCCCTGACCTT	2163
QY	1737	CCGGAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGCAGCAACACCGGCTCACTCT	1796
Db	2164	CCGGAGGAGCAGTGTGCTGCTACTAACACCCGACGAGACCTCTTCAAGAGCTTCCGAGG	2223
QY	1797	CGCGTGCATGGGTGGCCCAAGTACTCCGGGTGTCTCCCGGCAAGTGTGCAAGCTCAT	1856
Db	2224	GCCCATGAGCTGGGTTCCTCCTACACAGGCGTGGCCCCCAGGACAGTGTGCAACTCAC	2283
QY	1857	CTCGGAGGCAATGGCAGTGGCTACTTCTATGTCTGSCACCCCAAGTGTGGAGCGAC	1916
Db	2284	CTGCCAGCCCGGCGACTGGGCTACTACTATGTCTGGAGCCACGGTGTGTAGATGGAC	2343
QY	1917	GCTGTGCTCTCTGACTCCACTCGGTCTGTCTCAAGGCAAGTGCATCAAGGCTGGCTG	1976
Db	2344	CCCTGTTCCTCCCGGACAGCTCTCGGTCTGTCTCCAGGCGGATGCATPCCATGCTGGCTG	2403
QY	1977	TGATGGGAACCTGGGCTCCAAAGAAGATTCGACAAGTGTGGGTGTGTGGGGAGACAA	2036
Db	2404	TGATCGCATCATTTGGTCCCAAGAAGATTTTGACAAGTGCATGCTGTCCGAGGGGACGG	2463
QY	2037	TAAGAGCTGCAAGAAGTGTGACTGGACTCTTCAACAAAGCCCATGTGATGGCTTCAATTCGT	2096
Db	2464	TTCTGTTGACGACAGCAGTCAAGGCTCTTTCAGGAAATTCAGGTACGGATACAACAATGT	2523
QY	2097	GGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCCGCGCAGCGGTTTACAAGGCT	2156
Db	2524	GGTCACTATCCCGCGGGGCGCCACCATTTCTGTCCGGCAGCAGGAAACCCCTGGCCA	2583
QY	2157	GATCGGGGATGACAACTACCTGGCTCTGAAGAACAGCCCAAGTACCTGTCTCAAGCG	2216
Db	2584	CCGGAG-----CATCTACTTGGCCCTCAAGCTGCACAGTGGCTCTATGCCCTCAATGG	2637
QY	2217	GCATTTGCTGTGTGGGCTGGAGCGGACCTGGTGTG--AAGGGCAGTCTGCT--TGCG	2273
Db	2638	TGAATACACGCTGATSCCTCCCGCACAGATGTGTACTGCTGGGCGAGTCACTTGGC	2697
QY	2274	GTACAGCGCAGCGGACAGCGGTGGAGAGCTGTGAGGCTTCCCGGCGCCATCTCTGGAGCC	2333
Db	2698	CTACAGCGGGCCACTGACGCTCAGACACTGTTCAGGCCATGGCCAC--GGCCCAAGCC	2757
QY	2334	GCTGACCTGGAGTCTCTCCGTTGGGGAAGATGACACCGCCCGGGTCCCGCTATCTCCTT	2393
Db	2758	TTTGACACTGCAAGTCTTAGTGGTGGCAACCCCGAGGACACACGCTTCCGATACAGCTT	2817
QY	2394	CTATCTGCC 2403	
Db	2818	CTTGTGGCC 2827	

Search completed: May 2, 2003, 00:40:34
Job time : 393 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2003, 19:28:11 ; Search time 3890 Seconds
(without alignments)
11878.078 Million cell updates/sec

Title: US-10-009-332-2
Perfect score: 2853
Sequence: 1 atgttttctggcgcctcct.....gcgtctcaggccgtgctga 2853

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: em_estba:**
- 2: em_esthum:**
- 3: em_estin:**
- 4: em_estmu:**
- 5: em_estov:**
- 6: em_estpl:**
- 7: em_estro:**
- 8: em_htc:**
- 9: gb_est1:**
- 10: gb_est2:**
- 11: gb_htc:**
- 12: gb_est3:**
- 13: gb_est4:**
- 14: gb_est5:**
- 15: em_estfun:**
- 16: em_estom:**
- 17: gb_gss:**
- 18: em_gss_hum:**
- 19: em_gss_inv:**
- 20: em_gss_pln:**
- 21: em_gss_vrt:**
- 22: em_gss_fun:**
- 23: em_gss_man:**
- 24: em_gss_mus:**
- 25: em_gss_other:**
- 26: em_gss_pro:**
- 27: em_gss_fod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	634	22.2	2931	11	BC003269 Mus muscu
2	531.8	18.6	849	13	BI103177 602889417
3	512.6	18.0	681	13	BG921734 602825569
4	406	14.2	961	10	BB612189 BB612189
5	391.4	13.7	522	10	BE553572 ur47b04.y
6	349.6	12.3	544	12	BF078689 229090 MA

7	339.4	11.9	627	10	BE281680
8	270.6	9.5	654	10	BE016461
9	263.4	9.2	738	13	BI1819909
10	256.4	9.0	746	13	BI733795
11	245.8	8.6	661	10	BB220373
12	244.2	8.6	660	10	BB651049
13	242	8.5	662	10	BB642806
14	241.8	8.5	951	14	BQ30852
15	231.4	8.1	945	12	BF030017
16	229.4	8.0	575	13	BM570576
17	227.6	8.0	984	17	CNS0457M
18	226	7.9	797	14	BQ958765
19	222.6	7.8	305	12	BF320986
20	219.6	7.7	598	13	BM069515
21	215	7.5	674	10	BB225365
22	209.4	7.3	723	12	BG826611
23	209.2	7.3	949	14	BQ993675
24	206.2	7.2	652	9	AI400905
25	206	7.2	420	10	AW353152
26	205	7.2	681	12	BE912007
27	202.8	7.1	562	13	BQ009737
28	201.2	7.1	602	10	AW137467
29	200.2	7.0	601	14	BQ332355
30	196.4	6.9	680	12	BF446232
31	195.6	6.9	528	9	AA288689
32	190	6.7	518	17	AZ374992
33	188	6.6	689	12	BG242316
34	186.4	6.5	770	17	CNS03W7R
35	185.6	6.5	667	10	BG624639
36	185.4	6.5	504	12	BG553803
37	185.4	6.5	522	10	AW046509
38	184.8	6.5	640	12	BE742606
39	183.2	6.4	543	14	BQ011430
40	180.6	6.3	468	10	BE646878
41	180.6	6.3	475	9	AI148739
42	178.4	6.3	264	10	BB576723
43	177.4	6.2	566	10	AV617111
44	176.2	6.2	257	9	AI644268
45	172.4	6.0	464	10	BE666088

ALIGNMENTS

RESULT 1
BC003269 2931 bp mRNA linear HTC 07-AUG-2002
LOCUS Mus musculus, Similar to a disintegrin-like and metalloprotease
DEFINITION (repolysin type) with thrombospondin type 1 motif, 1, clone
IMAGE:3499206, mRNA.
ACCESSION BC003269
VERSION BC003269.1 GI:14707747
KEYWORDS HTC
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2931)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: angbcm@bcm.tmc.edu
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Toon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 9 Row: b Column: 7
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
1. .2931
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="IMAGE:3499206"
/tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

BASE COUNT 708 a 812 c 811 g 600 t

Query Match 22.2% Score 634; DB 11; Length 2931;
Best Local Similarity 61.0%; Pred. No. 1e-125;
Matches 1084; Conservative 0; Mismatches 680; Indels 12; Gaps 3;

```
QY 631 AAGGTTTCGTCTATCCCGCGTAGTGGAGACGCTGGTGTGCGGGACGAGTCAATG 690
D 1144 AAGCGATTGTGTCACGCCCGGTATGTGGAACCAATGCTCGTGGTGACCAAGTCCATG 1203
QY 691 GTCAAGTTCCACGGCGGGAGCTGGAAATATCTGTGACGCTGTGCGCAACGGCGGCG 750
D 1204 GCCGACTTCACGGCAGCGGTCTAAAGCAATTAACCTTAAACCTGTCTCGGTGGCAGCC 1263
QY 751 CGACTTACCGCCATCCAGATCTCAACCCATCAACATGCTGTGTCAGTGGTCAAGTCTG 810
D 1264 AGGTTTACAGCATCCGACCGCTTAGGAATTCATTAAGCTGTGTGGTGTGGAAGATCTTG 1323
QY 811 CTCTTAGAGATCTGTACTCGGGCCCAAGGTACCGGCAATCGGCGCTGACGCTGCGC 870
D 1324 GTCATATAGAGGACGAGAGGACAGAGTACCTTCCATGACGCTCTCACCTTCGG 1383
QY 871 AACTTCTGTGCTGGCAGAGAAGCTGAACAAAGTGTGATGACAAAGCACCCTGAGTACTGG 930
D 1384 AATTCTGCAACTGGCAGAAACACACACAGCCCGGATCCGGGATCCAGAGCACTAT 1443
QY 931 GACACTGCCATCTCTTACCAGCAGGACCTGTGTGGAGCCACCCTGTGACACCTG 990
D 1444 GACACTGCAATCTGTTCACAGACAGGATTTATGTGGCTCCACACAGCTGTGACACTCTC 1503
QY 991 GGCATGGCTGATGTGGGTACCATGTGTGACCCCAAGAGAAGCTCTGTGCTATTGAGGAC 1050
D 1504 GGGATGGCAGATGTGGAACTGTATGTGACCCAGCAGGAGCTCTCAGTCATAGAAGAT 1563
QY 1051 GATGGGCTTCCATCAGCCTTCCACACTGCCCCAGAGCTGGGCCACGTGTTCAACATGCC 1110
D 1564 GATGGTTTGAAGCGCGCTTCCACACAGCCCAAGTAATGGGCCATGTGTTTAACATCCG 1623
QY 1111 CATGACATGTGAAAGTGTGAGGAGGTGTTGGGAAGCTCCCGAGCCCAACCATCATG 1170
D 1624 CACGATGATGTAGACACTGTGCGCAGCTTGAATGGTGTGACTGGCGATTTCTATCTGATG 1683
QY 1171 TCCCGGACCTTCATCCAGATCGAGCTGCCAACCCCTGGTCAGCCTGCAGTGTGCCATC 1230
D 1684 GCCTGATGCTCTCCAGCTTAGACCATAGCAGCCCTGGTCACTTGCAGTGCCTACATG 1743
QY 1231 ATCACCGACTCTCTGGACAGCGGACGAGTGTGCTCTCTGGACCAACCCAGCAAGCC 1290
D 1744 GTCAGCTCTCTTCCATAGATAATGGACACGGGAATGTTTGTATGGACAGCCCGCAGAA 1803
QY 1291 ATCTCTCTGCCGAGGATCTCGCGGGCCGACGCTACACCTGTAGCCAGCTGCGAGCTG 1350
```

```
DB 1804 ATCAAGCTCCCTTCTGATCTTCCCGGTACCTTGTACGATGCCAACCGCAGTGTGAGTTT 1863
QY 1351 GCTTTTGGCGTGGGTCCCAAGCCCTGTCTTACATGCAGTAC---TGACCAAGAGTGTGG 1407
D 1864 ACATTTCGGAGAGGAATCCACGACATGCCCTGATGCAGCCAGCAGATGTACTACTCTGG 1923
QY 1408 TGCAACGGGAAGGCCAAGGACAGATGTTGCCAGAGCCGCCACTTCCC"TGCGCCGAT 1467
D 1924 TGCACTGGCACTCGGTGGTGTACTTGTGTGCCAAACAAAACACTTCCCTTGGCAGAT 1983
QY 1468 GGCACACGCTGTGGGAGGCAAGCTCTGCCTCAAGGGGCTGGTGGAGACACACAC 1527
D 1984 GGCACACGCTGTGGAGAAAGGAAGTGGTGTGTCTAGTGGCAAGTGGTGAACAAACAGAC 2043
QY 1528 CTCACCAAGCACAG-----GGTGTGATGTTCTTGGGCAAAATGGATTCCTATGGCCCC 1581
D 2044 ATGAAGCATTTTGTCTACTCTCTTTCATGGAAGCTGGGACCATGGGACCGTGGGAGAC 2103
QY 1582 TGCTCGCCACATGTGGTGGGGCGTGCAGCTGGCCAGGAGGAGTGCACCAACCCACC 1641
D 2104 TGCTCAAGAACCTGTGGTGGTGGAGTTCAATACACAATGAGAGAATGTGACAAACCCAGTC 2163
QY 1642 CCTGCCACGGGGCAAGTACTGCGAGGAGTGAAGTGAATACCGATCTCTGCAACCTG 1701
D 2164 CCAAGACCGGAGGAAGTACTGTGAAGCAACAGTCCGCTACAGTCTCTGTAAACATC 2223
QY 1702 GAGCCCTGCCCCAGCTCAGCCCTCCGGAAGAGCTTCCGGGAGGAGCAGTGTGAGGCTTTC 1761
D 2224 GAGGACTGTCCAGACAATAA---CGGAACAAAGCTTCAGAGAGGAGCAGTGCAGGCGCAC 2280
QY 1762 AACGGGTACACCAACAGCACCACCGGCTCACTCTCGCGCTGGCATGGTGGTCCCAAGTAC 1821
D 2281 AATGAGTTTTCACAAAGCTTCTTTGGGAATGAGCCCACTGTAGAGTGACACCCCAAGTAC 2340
QY 1822 TCCGGGTCTCTCCCGGGAGCAAGTGCAGCTCATCTGCGGAGCCAAATGGCACTGGCTAC 1881
D 2341 GCGGGGTCTCCGCAAGGACAGTGCAGCTCACTGTGAAAGCCAAAGGATTTGGCTAC 2400
QY 1882 TTCTATGTCTGGCACCACAAAGTGGTGGAGCGCAGCTGTGCTCTCTCTGACTCCACTCC 1941
D 2401 TTTTTCGTCTTACAGCCCAAGGTTGTAGTGGCACTCCCTGTAGTCCAGACTCTACCTCT 2460
QY 1942 GTCTGTGTCACAGGCAAGTGCATCAAGCTGGCTGTGTGATGGGAACCTGGGCTCCCAAGAG 2001
D 2461 GTCTGTGTCACAGGCAAGTGTGTGAAGCTGGCTGTGTGATCGATCATAGACTCCCAAGAG 2520
QY 2002 AGATTGCAACAGTGTGGGTGTGTGGGGAGACAATAAGAGCTGCAAGAGTGAAGTGAAG 2061
D 2521 AAGTTTGTAGTGTGGCGTTTGTGGAGGAAACGGTTCCACATGCAAGAAGATGTCAAG 2580
QY 2062 CTCTTCCACCAAGCCATGATGGCTACAAATTTCTGGTGGCCATCCCGCAGCGCCCTCA 2121
D 2581 ATAGTCACTAGTACAAAGACTGGGTATCATGATTTGTGCACAAATTCCTGTCTGAGCCACC 2640
QY 2122 AGCATCGACATCCCGCAGCGGTTACAAAGGCTGATCGGGGATGACAACTACCTGGCT 2181
D 2641 AACATTGAGTGAACATCGGAATCAAGGGGTCCAGAAACAATGGCAGCTTCTGGCT 2700
QY 2182 CTGAAGAACAGCCAAAGCAAGTACCTGTCAACGGGCAATTTCTGGTGTGGCGGTGGAG 2241
D 2701 ATTAGAGCGCTGTGTTACCTATATCTGAAATGGAACCTTCACTCTCTCCACACATAG 2760
QY 2242 CGGGACCTGGTGTGAAGGCACTGTCTGCGGTACAGCGGACGCGGACACGGTGGAG 2301
D 2761 CAAGACTCTCACTTACAAAGGTACTGTCTTAAGGTACAGTGGTTCCTGGCTCGGCTGGAG 2820
QY 2302 AGCCTCAGGCTTCCCGGCGCACTCTGGAGCGGCTGACCGTGGAGGCTCTCTCCGTGGGG 2361
D 2821 AGAATCCGAGCTTTAGTCCACTCAAGAAACCTTAAACCATCCAGGTTCTTATGTAGGC 2880
QY 2362 AAGATGACACCGCGCGGTTCGCTACTCTCTCTAT 2397
```

```

Db 2881 CATGCTCTCCGACCCAAAATAATTCACCTACTTT 2916
      BII03177      849 bp      mRNA      linear      EST 26-JUN-2001
      LOCUS      602889417f1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5044493
      DEFINITION      5', mRNA sequence.
      ACCESSION      BII03177
      VERSION      BII03177.1 GI:14554070
      KEYWORDS      EST.
      SOURCE      house mouse.
      ORGANISM      Mus musculus
      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
      1 (bases 1 to 849)
      NIH-MGC http://mgc.nci.nih.gov/
      National Institutes of Health, Mammalian Gene Collection (MGC)
      Unpublished (1999)
      Contact: Robert Strausberg, Ph.D.
      Email: cgapbs-r@mail.nih.gov
      Tissue Procurement: Jeffrey E. Green, M.D.
      CDNA Library Preparation: Life Technologies, Inc.
      CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: L1AM11122 row: a column: 06
      High quality sequence stop: 743.
      Location/Qualifiers
        1..849
          /organism="Mus musculus"
          /strain="FVB/N"
          /db_xref="taxon:10090"
          /clone="IMAGE:5044493"
          /clone_lib="NCI_CGAP_Kid14"
          /lab_host="DH10B (T1 phage-resistant)"
          /note="Organ: kidney; Vector: pCMV-SPORT6; Site.1: NotI;
          Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
          Average insert size 1.75 kb. Constructed by Life
          Technologies. Note: this is a NCI_CGAP Library. |"
      BASE COUNT      193 a 250 c 250 g 156 t
      ORIGIN
      Query Match      18.6%; Score 531.8; DB 13; Length 849;
      Best Local Similarity 85.1%; Pred. No. 6.6e-104;
      Matches 699; Conservative 0; Mismatches 112; Indels 10; Gaps 9;
      QY 1189 ATGAGCGGTGCGAAGCCCGTGGTACGCTGCGAGTGTGCGATCATCATCAC-CGACTTCTCTGG 1246
      Db 26 ATAGACCGGTGCCAACCCCATGGTGCAGCTGCAGCGCTGCCATTATAACGTGACTTCTCTGG 85
      QY 1247 ACAGCGGGCAGGTGACTGCTCTCGACCAACCCAGCAAGCCATCTCCCTGCCCGAGG 1306
      Db 86 ATAGTGGCATGTGACTGCTCTCGACCGAGCCAGCAAGCCATCATCCCTGCTGAGG 145
      QY 1307 ATCTCGGGCGCCAGCTACACCTGAGCGCAGTGCAGTGGCTTTTGGCTGGGCT 1366
      Db 146 ACCTCGGGGCACAAAGCTACAGTTTCAGCCAGCAGTGCAGCTGGCTTTGGGTGGGCT 205
      QY 1367 CCAAGCCCTGCTTACATGACGTACTGCACCAAGTGTGGTGCAACCGGGAAGCAAGG 1426
      Db 206 CTAAGCCCTGCGCATATATGCACTACTGTACAAAGCTGTGGTGCAAGGCAAGCAAGG 265
      QY 1427 GACAGATGTGTGCCAGACCGGCCTCTCCCTGGGCCATGGCACCAGCTGTGGCGAGG 1486
      Db 266 GGCAGATGTGTGCCAGACTCGCCACTTCCCTGGGACATGGCACCAGCTGTGGGGAAG 325
      QY 1487 GCAAGCTCTGCCCTCAAAGGGGCTGGTGGAGAGACACACCTCAACAGCAGCAGGGTGG 1546
      Db 326 GCAAGTTCTGCCCTCAAGGGAGCTGGTGGAGAGACACACCAACCAACAGTAGTACCGGTGG 385
      RESULT 3
      BG921734
      LOCUS      602825569f1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:4954456 5',
      DEFINITION      mRNA sequence.
      ACCESSION      BG921734
      VERSION      BG921734.1 GI:14302210
      KEYWORDS      EST.
      SOURCE      house mouse.
      ORGANISM      Mus musculus
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
      1 (bases 1 to 681)
      NIH-MGC http://mgc.nci.nih.gov/
      National Institutes of Health, Mammalian Gene Collection (MGC)
      Unpublished (1999)
      Contact: Robert Strausberg, Ph.D.
      Email: cgapbs-r@mail.nih.gov
      Tissue Procurement: Jeffrey Green M.D.
      CDNA Library Preparation: Life Technologies, Inc.
      CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      http://image.llnl.gov
      Plate: L1AM10915 row: i column: 17
      High quality sequence stop: 680.
      Location/Qualifiers
        1..681
          /organism="Mus musculus"
          /strain="FVB/N"
          /db_xref="taxon:10090"
          /clone="IMAGE:4954456"
          /clone_lib="NCI_CGAP_Mam6"
          /sex="female, virgin"
          /tissue_type="infiltrating ductal carcinoma"
          /dev_stage="5 months"
          /lab_host="DH10B"
          /note="Organ: mammary; Vector: pCMV-SPORT6; Site.1: SalI;
          Site.2: NotI; Cloned unidirectionally. Primer: Oligo dT.
  
```


Query Match 14.2%; Score 406; DB 10; Length 961;
Best Local Similarity 82.7%; Pred. No. 7.3e-77;
Matches 488; Conservative 0; Mismatches 100; Indels 2; Gaps 2;

QY 1 ATGCTTTTGTGGGATCTTAACCTTGGCTTTTCGCCGGCGGACCGCTGGAGGCTTTGAG 60
Db 321 ATGCTTTTGTGGGATCTTAACCTTGGCTTTTCGCCGGCGGACCGCTGGAGGCTTTGAG 60
QY 61 CCAGAGCGGGAGGTAGTCTGCTCCATCCGACTGGACCGGACATTAACCGCGCGCTAC 120
Db 381 CCAGAGTGGAGGTAGTCTGCTCCATCCGAGCGGACCGGACATTAACCGCGCGCTAC 440
QY 121 TACTGGGGGCTCCCGAGGACTCCGGGATCAGGGACTCATTTTTCAGATCACAGCATTT 180
Db 441 TACAGGAGGCTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCCGCTTCC 240
QY 181 CAGGAGGCTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCCGCTTCC 240
Db 501 CAGGAGGCTTTTATCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCCGCTTCC 240
QY 241 ACTGAGCATCTGGGCGCTCCCGCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGGAGCG 300
Db 561 ACTGAGTATAGTGTCCCGCTGACAGGCTCAGTGGAGCTCTCTAGACCTGGAGCG 620
QY 301 TGTCTTATTTCTGGGAGCTGAACCGCGGAGCTCCTGCTGCTGTGAGCTGTGC 360
Db 621 TGTCTTATTTCTGGGAGCTGAACCGCGGAGCTCCTGCTGCTGTGAGCTGTGC 679
QY 361 GGGGGCTCCCGGAGCTTTGGCTACCGAGGCGGAGTATGCTATAGCCGCTGCCC 420
Db 680 GGGGGCTCCCGGAGCTTTGGCTACCGAGGCGGAGTATGCTATAGCCGCTGCCC 739
QY 421 AATGCTAGCGCGCGCGGCGAGCGACAGCGAGGCGGACACCTTCTCCAGCGCGCG 480
Db 740 AACACAGCGCGCGAGGCGGAGCGGCGACAGCGAGGCGGACACCTTCTCCAGCGCGG 799
QY 481 GGTGTTCCGGGGGCTTCCGGAGACCGACCTCTCGTGGGGGTGGCTCGGGCTGG 540
Db 800 GGTGCTCTGTATGGCTTCCGGAGACACACCTCTCGTGGGGGTGGCTCGGGCTGG 858
QY 541 AACCCCGCATCTACGGCGCTTGGACCTTACAGCGCGCGGCGGCGG 590
Db 859 TACCCCGCATCTAGGCTCTCGACCCCATAGCCACCGAGGCGCGG 908

RESULT 5
BE553572
LOCUS
DEFINITION
ur47b04.y1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:3153391 5'
similar to TR:054768 054768 A DISINTEGRIN AND METALLOPROTEASE WITH
THROMBOSPONDIN MOTIFS ; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
TITLE
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1056147
Seq primer: -40RP from Gibco
High quality sequence stop: 385.
Location/Qualifiers
1..522
/organism="Mus musculus"

/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:3153391"
/clone_lib="NCI_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo df.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT 116 a 133 c 158 g 115 t
ORIGIN

Query Match 13.7%; Score 391.4; DB 10; Length 522;
Best Local Similarity 86.7%; Pred. No. 8.4e-74;
Matches 431; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1888 GTGCTGGCACCACCAAGTGTGGAGCGACGCTGTCTCTCTGACTCCACCTCGCTCTGT 1947
Db 19 GTCCGAGCACCTAAGTGTGGATGGTACGCTGTGTACTCTGACTCCACCTCGCTCTGT 78
QY 1948 GTCCAAAGCAAGTGCATCAAGGCTGGGCTGTGATGGAACTGGGCTCCAAAGACAGATTC 2007
Db 79 GTCCAAAGCAAGTGCATCAAGGCTGGGCTGGCAGCGGAATCTGGGCTCCAAAGAA/AAATTT 138
QY 2008 GACAAAGTGTGGGCTGTGGGGGAGACAATAAGAGCTCAAGAGAGTGAAGAGGTTGAC/AGACTCTTC 2067
Db 139 GACAAAGTGTGGGCTGTGGGGGAGACAATAAGAGCTCAAGAGAGTGAAGAGGTTGAC/AGACTCTTC 198
QY 2068 ACCAAGCCCATGCTATGCTTACAAATTTCTGTGTGGCCATCCCGCAGGCGCTCAAGCATC 2127
Db 199 ACCAAGCCCATGCTATGCTTACAAATTTCTGTGTGGCCATCCCGCAGGCGCTCAAGCATC 258
QY 2128 GACATCCCGAGCGGCTTACAAAGGCTGTATCGGGATGACAACTACCTGGGCTCTGAAG 2187
Db 259 GATATCCCTGACGCTGTGTACAAAGGCTCATTTGAGATGACAACTACCTGGGCTCTGAAG 318
QY 2188 AACAGCCAAAGCAAGTACCTGCTCAAGCGGCAATTTCTGTGTGGCTGGGCTGGAGCGGAC 2247
Db 319 AACAGCCAAAGCAAGTACCTGCTCAAGCGGCAATTTCTGTGTGGCTGGGCTGGAGCGGAC 378
QY 2248 CTGTTGTTGAAGGCGAGTCTGCTCGGTPACAGCGGCAAGCGGCTGGAGAGCGCTG 2307
Db 379 CTGTTGTTGAAGGCGAGTCTGCTCGGTPACAGCGGCAATTTCTGTGTGGCTGGAGAGCGCTG 438
QY 2308 CAGGCTTCCCGCGGCTTCCGAGCGGCTGACCGTGGAGGCTCTCTCGGTTGGGGAAGATG 2367
Db 439 CATGCTTATCGACCCACTACTCGAGGCGACTCACCTGGAGGTCCTGTCTCGGTTGGGGAAGATG 498
QY 2368 ACACCGCCCGCGGCTCCG 2384
Db 499 ACACCGCTCTGTGGCG 515

RESULT 6
BF078689/c
LOCUS
DEFINITION
229090 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Sus scrofa
pig.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
AUTHORS
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Bennett, G.E., Laegreid, W.W.
and Keeler, J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)

COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single-pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGTCAGGAGC
Plate: 43 row: P column: 17
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1..544
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DHL0B"
/note="Vector: PCMV SPORT6; Site.1: NotI; Site.2: SalI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 117 a 154 c 169 g 104 t
ORIGIN
Query Match 12.3%; Score 349.6; DB 12; Length 544;
Best Local Similarity 88.6%; Pred. No. 8.1e-65;
Matches 379; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 559 GCCTTGACCTTACAAAGCGGGGGGGCTTCGGGAGAGTGTAGCGGGCGGAGG 618
DB 544 GCCTTGACCTTACAAAGCTTCGGGATAGGCTTAGGGAGGTCGACGGCGGAGG 485
QY 619 TCTGGGGCGGCCAAGCGTTTCGTGTCTATCCCGGTAGCTGGAGAGCGTGTGTGCGG 678
DB 484 TCCGGGGCGGCCAAGCGTTCGTGTCTATCCCGGTAGCTGGAGAGCGTGTGTGCGA 425
QY 679 GACAGTCAATGTCAAGTTCACAGCGCGGACCTCCGACATATTATCTGTGACGCTGTG 738
DB 424 GACAGTCCATGTCAAGTTCACAGCGCGGACTTGGAGCATATTCTGTGACGCTGTG 365
QY 739 GCAACGGCGCGGACTCTACCGCCATCCAGCATCCTCAACCCCATCAACATCGTTGTG 798
DB 364 GCCACGGCGCGGACTCTACCGCCATCCAGCATCCTCAACCCCATCAACATCGTGTG 305
QY 799 GTCAAGGTGCTCTTTAGAGATCTGTACTCCGGGCCCAAGGTCAACGGCAATCGGCC 858
DB 304 GTCAAGGTGCTCTTTAGAGATCTGTACTCCGGGCCCAAGGTCAACGGCAATCGGCC 245
QY 859 CTGACGCTGGCAACTTCTGTGCTGGCAGAGAGCTGAACAAAGTGAAGTGAACAGCAC 918
DB 244 TTGACGCTGGCAACTTCTGTGCTGGCAGAGAGCTGAACAAAGTGAAGTGAACAGCAT 185
QY 919 CCGGAGTACTGGGACACTGCCATCTCTTACAGCAGGAGACTGTGTGGAGCCACCC 978
DB 184 CTTGAGTACTGGGACACTGCCATCTCTTACAGCAGGAGACTGTGTGGAGCCACCC 125
QY 979 TGTGACAC 986
DB 124 TTGGACCC 117
RESULT 7
LOCUS BE281680
DEFINITION 601099657F1 NC1_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3491991 5',
mRNA sequence.
ACCESSION BE281680
VERSION BE281680.1 GI:9156727
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 627)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM8537 row: a column: 16
High quality sequence stop: 530.
Location/Qualifiers
1..627
/organism="Mus musculus"
/strain="Czech II"
/db_xref="taxon:10090"
/clone_image="3491991"
/clone_lib="NC1_CGAP_Lu29"
/tissue_type="spontaneous tumor, metastatic to mammary."
Stem cell origin.
/lab_host="DHL0B"
/note="Organ: lung; Vector: PCMV-SPORT6; Site.1: SalI;
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 153 a 159 c 177 g 138 t
ORIGIN
Query Match 11.9%; Score 339.4; DB 10; Length 627;
Best Local Similarity 83.6%; Pred. No. 1.3e-62;
Matches 504; Conservative 0; Mismatches 86; Indels 13; Gaps 10;
QY 1850 AGCTCATCTGCGGAGCCAAATGGCCTACTTCTATGTG-CTGGCACCACCAAGTGGTG 1908
DB 1 AGCTAATCTCGGAGCCAAATGGGACTGGCTACTTCTATGAAGTGAAGTGGTG 60
QY 1909 GAGGCGAGCTGTGCTTCCTGAC-TCCACCTCGTGTGTGTCCAAAGCAAGTGCATCAA 1967
DB 61 GATGTAGCTGTGTACTCTCTGACATCCACCTCGTGTGTGTCCAAAGCAAGTGCATCAA 120
QY 1968 -GGCTGGCTGTGTTGGGAACCTGG-GCTCCAAAGAGAGATTTCGACAA-GTGTGGGTGTG 2024
DB 121 TGGCTGGCTGGCGGGAATCTGGTGTCTCCAAAGAGAAATTTGACACATGTGTGTGTG 180
QY 2025 TGGGAGACACAATAA-GAGCTGCAAGAGGTGACTGGACTCTTCCACCAAGCCATGCATG 2083
DB 181 TGGTGGACACAATAATGAGCTGAAGAGGTGACAGGACTCTTCCACCAAGCCTTGCACG 240
QY 2084 GCTACAATTCGTGTGGCCATCCCGCAGCGGCTCAAGCATGCATCGCCAGCGCG 2143
DB 241 GCTACAATTCGTGTGGCCATCCCGTGTGTGGCCCTCCAGCATGTATTCGTGAGCGTG 300
QY 2144 GTTACAAAGGCTGATCGGGATGACACTACCTGGCTCTGAAGAACAGCCAAAGCAAGT 2203
DB 301 GTTACAAAGGCTCATTTGGAGATGACAACTACCTGGCTTTGAAGAACAGCCAAAGCAAT 360
QY 2204 ACCTGCTCAAGGGCATTTTCGTGTGTGTCGGCGGTGGAGCGGACCTGGTGTGAA--GGG 2261
DB 361 ACCTGCTCAATGGGCACTTACTGTGTATCCGCTGTAGAGCGGACCTGTGTGTAACAGG 420
QY 2262 CAGTGTGCTGGGTACACGGCAGCGGACAGCGGTGGAGAGCTGCAGGCTTCCCGGCC 2321
DB 421 CAGTGTGCTAGCTATAGTGGCACTGTGTGCTGCTGAGTGGAGAGCTGCAGGCTTCTC-GAC 479
QY 2322 CATCTGAGCGGCTGACCGTGGAGGTCTCTCTCGTGGGGAAGATGACACCGCCCGCGGT 2381

D	B	480	CATCTCGGAGGCACTCACGGTGGAGGTCTTGACGGTGGGGGAAGTAGCACCCGACC ---GT	536
Q	Y	2382	CGCGTACTCCTCTTCTATCTGCCCAAAGACCTCGGAGGACAAGTCTCTCATCCCAAGA	2441
D	B	537	GTGCGTATCTCTACTACTTGGCCAAAGGACCTC-GGAGGACAAGTCCATCGCCCCCAAGAC	595
Q	Y	2442	CCC CTC 2444	
D	B	596	CTC 598	
R	E	S	U	L
RESULT	8			
BE016461				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes				
; Cyprinidae; Danio.				
1 (bases 1 to 654)				
Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy				
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood				
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,				
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,				
Kohn,S., Shin,T., Jackson,I., Cardenas,M., McCann,R., Waterston,R.				
and Wilson,R.				
WashU Zebrafish EST Project 1998				
Unpublished (1998)				
Contact: Stephen L. Johnson				
Washington University School of Medicine				
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA				
Tel: 314 286 1800				
Fax: 314 286 1810				
Email: zbrafish@watson.wustl.edu				
cDNA Library Preparation: Ning Wu. cDNA Library Arrayed by:				
Research Genetics. DNA Sequencing by: Washington University Genome				
Sequencing Center Clone distribution: Research Genetics web				
address: http://www.researchgenetics.com/				
Seq primer: T7 ET from Amersham				
High quality sequence stop: 489.				
FEATURES				
source				
1..654				
/organism="Danio rerio"				
/db_xref="taxon:7955"				
/clone_lib="Zebrafish Research Genetics C32 fin"				
/tissue_type="Fin"				
/lab_host="GeneHops (HS996), a phage-resistant isolate of				
DH10B."				
/note="Vector: pT73D-Pac with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was prepared from zebrafish(C32) fin, and was then primed with a Not I - oligo(GT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is non-normalized. Library constructed by Ning Wu. NOTE:Clones from this library are only available thru Research Genetics (www.resgen.com)."				
BASE COUNT				
ORIGIN				
121 a				
193 c				
205 g				
135 t				
Query Match				
Best Local Similarity				
Matches 414; Conservative				
9.5%; Score 270.6; DB 10; Length 654;				
63.4%; Pred.No. 8e-48;				
0; Mismatches 239; Indels 0; Gaps 0;				
Q	Y	853	GCGCGCTGACGCTGCCAACACTTCTGTGCTTGGCAGAGAGCTGAACAAAGTAGTGAC	912
D	B	1	GCGCGCGGACGCTCAAGACCTTTCTGCAGTGGCAGAACCGACGACCTCTGTGATGAC	60

Db 418 GCTACTTTTTCGTCTTACAGCCCAAGTTGTAGATGGCACTCCCTGTAGTCAGACTCTA 477

QY 1937 CTCTCCGTCTGTGTCCCAAGCAAGTCATCAAGCGTGGCTGTGATGGACCTGGGCTCCA 1996
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 478 CCTCTGTCTGTGCAAGGGCAGTGTGTGAAGCTGGCTGTGATCGCATCATAGACTCCA 537
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1997 AGAGAGATTCGACAAAGTGTGGGTGTGGGGGAGACAATAAGAGCTGCAAGAGGTCA 2056
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 538 AAAAGAAGTTGGTAAGTGTGGCGTTGTGGAGAAACGGTTCCACATGCAAGAGATGT 597
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2057 CTGGACTCTTACCAGCCCATGCTGCTCAATTTCTGGTGGCCATCCCGCAGCG 2116
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 598 CAGGATAGTCTACTAGTACAGACCTGGGTATCATGACATGGTCCAAATTCCTGCTGAG 657
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2117 CCTCAAGCATGACATCCCGCAGCGGTTACAAAGGCTGATCGGGATGACAACTACC 2176
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 658 CCACCAACATTCGAATGAAACATCGGAATCAAGGGGTCCAGAAACAATGGCAGCTTTC 717
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 2177 TGGCTCTGA 2185
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 718 TGGCTATTA 726
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11

LOCUS BB220373

DEFINITION BB220373 RIKEN full-length enriched, adult male aorta and vein Mus musculus cDNA clone A530059019 3' similar to AF140673 Mus musculus putative secreted metalloprotease ADAMTS5 (Adams5) mRNA, mRNA

ACCESSION BB220373

VERSION BB220373.2 GI:16333495

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 661)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okada,Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT On Jun 30, 2000 this sequence version replaced gi:8885326.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,K., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES

Location/Qualifiers

source

1..661
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A530059019"
/clone_lib="RIKEN full-length enriched, adult male aorta and vein"
/sex="male"
/tissue_type="aorta and vein"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Site_1: Sali; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGATTAAATTAATATCCCTCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FUC 1."

BASE COUNT 160 a 167 c 180 g 154 t

ORIGIN

Query Match 8.6%; Score 245.8; DB 10; Length 661;
Best Local Similarity 60.9%; Pred. No. 1.7e-42;
Matches 400; Conservative 0; Mismatches 257; Indels 0; Gaps 0;

QY 861 GAGCTGCGCAACTTCTGTGCGCAGAAAGCTGAACAAGTGAACAAGCACC 920
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 4 GACCTCAAGAACTTTTGCAAATGGCAGCAACCAATACCACTAGGGATGATCAG 63
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 921 CGAGTACTGGACATGCCATCCTCTTACCAGGAGGACCTGTGTGGAGCCACCCTG 980
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 64 AGAGCTACTAGATGCGACCATCTGTTCACCGAGAGGATTTATGTGGCATCATCATG 123
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 981 TGACACCTCTGGCATGGCTGTGGGTACCATGTGTGACCCCAAGAGAGAGTCTCTGT 1040
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 124 TGACACCTCTGGCAATGGCAGAGCTTGGGACCATATGTTCTCGGAGCGAGCTGTGAGT 183
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1041 CATTGAGGAGATGGCTTCCATACGCTTACCCTGCCACGAGCTGGGCCCATGTGT 1100
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 184 GATTGAAGATGATGGCTCCATGCGACCTTCACTGTGGCTCATGAATTTGGCTCTACT 243
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1101 CAACATGCCCATGACAAATGTGAAAGTCTGTGAGGAGGTGTTTGGGAAGCTCCGAGCAA 1160
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 244 TGGCTTCTCTAAGAGATTCAAATCTGTGAAGAGAACTTCGGTACTACAGAGACAA 303
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1161 CCACATGATGTCCCGACCTCATCCAGATCGACCGTCCCAACCTGGTGTGAGTGTGAG 1220
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 304 GCGTTTAATGTCTTCAATCTTTACCAGGATCGATCATCCAGCCCTGTGTCCAATGAC 363
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1221 TGCTGCCATCATCACCAGCTCTCTGGAGCGGGACGGTGACTGCTCTCTGGA:CAACC 1280

QY	1318	GCCAGCTACACCTTGAGGCACACAGTGCAGACTGGCTTTTGGCGTGGGCTCCAAAGCCCTGT	1377
Db	545	CAGACCTACGATGCCACCGACGTGAACCTTGACATTTGGCGCTGAGTACTCGTGTC	604
QY	1378	CCTTACATGCAGTACTGACCAAGCTGTGTGACCGGGGAAGGCCAAGG	1426
Db	605	CCTGGCATGATGATCTGTGCGCGGCTGTGTGTCTGTGTGTCGCCAAG	653

RESULT 13					
BB642806					
LOCUS	BB642806	662 bp	mRNA	linear	EST 31-AUG-2001
DEFINITION	BB642806	RIKEN full-length enriched,	adult	retina	<i>Mus musculus</i> cdNA
		clone A930039D02 5',	RNA	sequence.	

VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE
BB042806.1	EST.	house mouse	Mus musculus	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 662)

TITLE RIKEN MOUSE ESTS (Arakawa, T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Havaehizaki

Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

email: genome-resgsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
November 14, 2003

genes. *Genome Res.* 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details

```

FEATURES
source
Location/Qualifiers
1. 662
/organism="Mus musculus"
/db xref="taxon:10090"

```

```
/clone_lib="RIKEN full-length enriched, adult retina"  
/tissue_type="retina"  
/dev_stage="adult"
```

QY 1318 GCCAGCTACCCGTGAGCCAGCAGTGGCTTTGGCGTGCCCAAGCCCTGT, 1377
| | | | | | | | | | | | | | | | | | | |
Db 545 CAGACTACGATGCCACCAGCAGTGCACATTGGCGCTGAGTACTCGGTGTGC 604

Db
605 CCTGGCARGATGCTGTGGCCGCTGTGTGCTGTGTGGCCAGG 653

BB642806
LOCUS
DEFINITION
BB642806 662 bp mRNA linear EST 31-AUG-2001
BB642806 RIKEN full-length enriched, adult retina Mus musculus cdna
clone A930039D02 5' mRNA sequence.

Accession	Keywords	Source	Organism
00042000.1	EST.	house mouse.	Mus musculus
0113402028			Eubacterium

[illegible]

TITLE RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

CONTACT: roshunide hayasizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Sukenaka
Wako-shi, Saitama 351-0192, Japan

Fax: 81-45-503-9216
Email: genome-res@gsic.riken.go.jp,
URL: <http://genome.gsic.riken.go.jp/>
Carninci, P., Shibata, Y., Hayatsu, N.,
Sugabara, Y., Shibata, Y., Tachibana, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

R1EN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10.1101/275175

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a non-redundant cDNA library. *Genome Biol.* 11: 231-242 (2010).

K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., Arakawa, T., Ishii, Y. and Hayashizaki, Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. *J. Struct. Func. Genomics* 2 pre. L72-L86 (2001)

```

FEATURES
source
C mouse tissues.
location/Qualifiers
1..662
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A930039D02"
/clone_lib="RIKEN full-length enriched, adult retina"
/tissue_type="retina"
/dev_stage="adult"

```

/lab_host="DH10B"
/note=Site1: Salt; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trisnase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5'GAGAGAGAGATTCGAGTTCGATTAATTAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. -retina RNA was provided by Stefano Gustinich, Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA02115, USA, whose assistance we gratefully acknowledge."

8.5%;	Score 242;	DB 10;	Length 662;
ty 60.5%;	Pred. No. 1.le-41;		
ervative	0;	Mismatches 260;	Indels 0; Gaps 0;
CCATCGTTGGTCAAGTGGCTCTCTTAGAGATCGTGA	CTCGGCC	837	
CCCTGGCGGTGGTGAAGTGGTGGTCTGACGCGAC	AGGACGAGTCTG	64	

CAATGGGGCCCTGACGCTGGCGCAACTCTGTGCGCTGGCGCAAGAAGCTG 897
GAATGGGGCCAGCACCTCAGAACTTTGCAATGGCGACCAACAT 124

GGGATGATCAGGAAGACACTACGATCGAGCCATCTGTTCACCCGAGAG 184
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TAGCCACCACCTGTGCACCCTGGCGCATGGGTGATGTGGGTAACCATGTGT 1017

AAGCTGCTCTGTGCATTCAGGACGATGGGTTCCATCAGCCTTCAACC ACT 1077
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CAGTGTCAGTGATTGAAGATGATGCCCTCCATCGACGCTTCACTGTG 304

GGGCCAGGIGITTCACCAATGCCCATCAGACAATGTGAAGAAGTCTGTGAGGAG 1137
| | | | |
TGCGGCATCTACTTTGGCCCTTCTCATGACGATTCCAAATTTCTGTGAAGAG 364

TACAGAGACAAGCGTTAATGTCTTCAATCCTTACCAGCATCGATGCA 424
GTGAGCGTCGACTGTCGCCATCATCCGACTTCCTGGACAGCGGGCAC 1257

CCTGGACCAACCCAGCAGCCCATCTCCCTGCCGAGGATCTGCCGGC 1317
 || |||| ||| | |||| | | |||| |||| |||| |||| |||| ||||
 GCTPAGCCTACACGGAAGCAGATTTTGGCTCCGAGGAACTCCCGAGA 544

CCTGAGCAGCAGTGGCGCTTTTGCGGTGCCAAGCCCTGT 1377
||||| ||||| | | ||||| | | ||
TGCCCCCAGCAGTGCACACTTTGGGCCTGAGTACTCGGTGTC 604

TGTCTGTGCGGGCTGGTGTGCTGTGGTGCCTCCAGGCCCAATTG 662

```
RESULT 14
BQ930852
LOCUS
DEFINITION BQ930852 951 bp mRNA linear EST 20-AUG-2002
IMAGE:6395437 5', mRNA sequence.
ACCESSION BQ930852
VERSION BQ930852.1 GI:22345883
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL (bases 1 to 951)
COMMENT NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: the Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13890 row: b column: 14
High quality sequence stop: 644.
FEATURES
source
1. .951
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_lib="IMAGE:6395437"
/lab_host="NCI_CGAP_Co24"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 189 a 143 c 180 g 135 t
ORIGIN
Query Match 8.5%; Score 241.8; DB 14; Length 951;
Best Local Similarity 62.2%; Pred. No. 1.4e-41;
Matches 397; Conservative 0; Mismatches 238; Indels 3; Gaps 1;
Qy 1549 GGTTCCTGGGCAATGGGATCCCTATGGCCCTGCTCGCGACATCTGTGGGGGGCGTG 1608
Db 1 GGAAGCTGGGACCATGGGACCGTGGGAGACTGCTCAAGAACCTGTGTGGTGGAGTT 60
Qy 1609 CAGCTGGCAGGAGCAGTGACCAACCCACCCCTGCCAACGGGGCAAGTACTGCGAG 1668
Db 61 CAATACACATGAGAGATGTGCAACCCCAAGTCCCAAGAACGGAGGAAGTACTGTGAA 120
Qy 1669 GGAGTGAGGTGAATACCGATCTCGAACCTGGAGCCCTGCCACAGCTCAGCCTCGGA 1728
Db 121 GGCAACGAGTCCGCTACAGTCTCTGTAACTCGAGGACTGTCCAGAC---AATAACGGA 177
Qy 1729 AAGAGCTTCCGGGAGGAGCAGTGTGAGCTTTCAACGGCTTACACACAGCAACCAACCG 1788
Db 178 AAAAGTTTCAGAGAGGAGCAGTGGCGGCGCACAAATGAGTTTCCAAAGCTTCTTTGGG 237
Qy 1789 CTCACCTCGCGGTGGATGGTGCCCAAGTACTCGGCGGTGTCCTCCCGGAGCAAGTGC 1848
Db 238 AATGAGCCCATGTAGAGTGACACACCAAGTACCGCGGCTCTCGCCAAAGGACAGTGC 297
Qy 1849 AAGCTCATCTCCGAGGCAATGGCAGTGGCTACTTCTATGTGCTGGCACCCCAAGGTGGT 1908
Db 298 AAGCTCACTGTGAAGCAAGGATTTGGCTACTTTTTCGTTCTTACAGCCCAAGTTGTA 357
Qy 1909 GACGGCAGCTGTGCTCTCCTGACTCCACCTCGCTGTGTCTTCCAAAGCAAGTGCATCAAG 1968
Db 358 GATGSCACTCCCTGTAGTCCAGACTTACCTCTGCTGTGTGTGTAAGGCGAGTGTGTA 417
```

```
Qy 1969 GCTGCTGTGATGGGAACCTGGGCTCCCAAGAGAGATTGACAAAGTGTGGGGTGTGGG 2028
Db 418 GCTGCTGTGATGGCATATAGACTCCAAAAGAGTTTGATAAGTGTGGCGTTGTGGGA 477
Qy 2029 GGAGACAATAAGAGCTCAAGAAGGTGACTGGACTTTCACCAAGCCCATCATGGGTAC 2088
Db 478 GGAACCGTTCCACATGCAAGAAGATGTCAGGAATAGTACATAGTACAAAGACCTGNGTAT 537
Qy 2089 AATTTCGTGGTGGCCATCCCGCAGCGCTCAAGCATCGACATCGGCCAGCGGGTTAC 2148
Db 538 CATGACATGTCACAAATTCCTGCTGGAGCCCAACATGAAGTGAACATCGGAATCAA 597
Qy 2149 AAAGGGCTGATCGGGGATGACAACTACCTGGCTGAA 2186
Db 598 AGGGGGTCCAGAAACAATGGCAGCTTCTGGCTATTAA 635
RESULT 15
BQ033017
LOCUS
DEFINITION BQ033017 945 bp mRNA linear EST 20-OCT-2000
IMAGE:601455701P1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3859104 5',
mRNA sequence.
ACCESSION BQ033017
VERSION BQ033017.1 GI:10740729
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: DCTD/DP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9592 row: b column: 01
High quality sequence stop: 642.
FEATURES
source
1. .945
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3859104"
/clone_lib="NIH_MGC_66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 220 a 246 c 326 g 153 t
ORIGIN
Query Match 8.1%; Score 231.4; DB 12; Length 945;
Best Local Similarity 61.8%; Pred. No. 2.3e-39;
Matches 422; Conservative 0; Mismatches 251; Indels 10; Gaps 3;
Qy 1534 AACACAGGCTGATGGTTCCTGGGCCAAATGGATGCCCTATGGCCCTGCTCGGCGACA 1593
Db 16 AGCCCCGTGTAGATGAGGCTGGGACCGCTGGGGACCTCGGGAGAAATGTTCTCGGACC 75
Qy 1594 TGTGTGGGGCGTGCAGCTGGCCAGGAGCAGTGCACCAACCCACCTCGCCACAGGG 1653
Db 76 TGTGGAGGAGGAGTACAGTTTTTTCACACCGCTGAGTGAAGGACCCCGAGCCTCAGAATGA 135
Qy 1654 GGCAAGTACTGCGAGGAGTGGGGTGAATACCGGATCTTGCACACCTGGAGCCCTGCC 1713
```

```

Db 136 GGAAGATACTGCCCTGGCTCGGAGAGCAAGTACCACTCATGCCACACGGAGGAATGCCGC 195
Qy 1714 AGCTCAGCCTCCGGAAGAGCTTCCGGAGGAGCAGTGTGAGGCTTTCACCGCTACAC 1773
Db 196 -----CCTGACGGGAAAGCTTCAGGGAGCAGTGTGAGAAGTATAATGCCCTACAA 249
Qy 1774 CACAGCACCAACCGCTTACTCTCGCCGTCGATGGGTGCCCAAGTACTCCGGCGGTGTCT 1833
Db 250 TACACTGACATGGACGGGAATCT---CCTGACGTGGGTCCCAAGTATGCTGGGTGTCC 306
Qy 1834 CCCCAGGACAGTGCAGCTCATCTCCGAGCCAATGGCACTGGCTACTTCTATGTGTG 1893
Db 307 CCCCAGGACCGCTCAAGTGTCTGCGAGCCCGGGAGGAGCGAGTCAAGGTGTC 366
Qy 1894 GCACCCAAAGTGTGGACGCGCTGTCTCTCTGACTCCACCTCCGCTCTGTGTCCAA 1953
Db 367 GAGGCCAAGGTGATTGATGCACCCCTGTGTGGCCAGAAACACTGGCCATCTGTGTCCGT 426
Qy 1954 GGCAAGTGCATCAAGGCTGCTGTGATGGGAACCTGGGCTCCAAAGAGATTCGACAA 2013
Db 427 GGCCAGTGTGTCAAGGCCGCTGTGACCATGTGGTGGACTCGCTCGGAAGCTGGACAAA 486
Qy 2014 TGTGGGTGTGTGGGGGAGACAATAAGAGCTGCAAGAAGTGAAGTCTTTCACCAAG 2073
Db 487 TCGGGGTGTGTGGGGCAAGGCAACTCTCGAGGAGTCTCCGGGTCCCTCACCCCC 546
Qy 2074 CCCATGCATGGCTACAATTTCTGGTGGCCATCCCGCAGCGCCCTCAAGCATCGACATC 2133
Db 547 ACCAATTATGGCTACAATGACATTTGCACCATCCAGCTGGTGGCCTAATATTGACGTG 606
Qy 2134 CGCCAGCGCGTTCACAAAGGCTGATCGGGATGACAACTACCTGCTCTGAAGAACAGC 2193
Db 607 AAGCAGCG-GAGCACCGGGTGTGCAGAACGATGGGACTACCTGGGGCTGAAGACGGT 665
Qy 2194 CAAGGCAAGTACCTGCTCAACGG 2216
Db 666 GATGGCAGTACCTGGTCAAGGG 688

```

Search completed: May 1, 2003, 22:50:30
Job time : 3933 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 00:40:42 ; Search time 7405 Seconds
(without alignments)
11212.735 Million cell updates/sec

Title: us-10-009-332-2

Perfect score: 2853

Sequence: 1 atcttttctggcattct.....gcgtctgagccgtgtgta; 2853

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size: 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rod:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2746	96.2	2853	9	HSA315733	AJ315733 Homo sapi
2	2089	73.2	2930	6	AX342635	AX342635 Sequence
3	2089	73.2	2937	6	AX319860	AX319860 Sequence
4	902	31.6	157963	9	AP002986	AP002986 Homo sapi
5	902	31.6	172905	2	AC025130	AC025130 Homo sapi
6	803	28.1	170682	2	AC023429	AC023429 Homo sapi
7	775	27.2	28000	9	AP003459	AP003459 Homo sapi
8	358	12.5	172905	2	AC025130	AC025130 Homo sapi
9	288	10.1	170682	2	AC023429	AC023429 Homo sapi
10	74	2.6	182656	2	AC101990	AC101990 Mus muscu
11	61	2.1	178764	2	AC126507	AC126507 Rattus no
12	44	1.5	178764	2	AC126507	AC126507 Rattus no
13	36	1.3	182656	2	AC101990	AC101990 Mus muscu
14	35	1.2	3927	10	BC009667	BC009667 Mus muscu
15	23	0.8	86229	2	AC099229	AC099229 Rattus no
16	23	0.8	176054	2	AC022377	AC022377 Homo sapi
17	23	0.8	178548	9	AC018831	AC018831 Homo sapi
18	23	0.8	198597	9	AC026218	AC026218 Homo sapi
19	23	0.8	202844	9	AC090956	AC090956 Homo sapi
20	23	0.8	229402	9	AC069271	AC069271 Homo sapi
21	22	0.8	132131	9	AL139159	AL139159 Human DNA
22	22	0.8	172913	2	AC124158	AC124158 Macaca mu
23	21	0.7	1333	5	AF070952	AF070952 Xenopus l
24	21	0.7	1437	5	AF070953	AF070953 Xenopus l
25	21	0.7	1456	5	AF072582	AF072582 Xenopus l
26	21	0.7	1979	8	HVACXP113	X78877 Hordeum vul
27	21	0.7	2290	10	BC006797	BC006797 Mus muscu
28	21	0.7	3315	9	BC016649	BC016649 Homo sapi
29	21	0.7	7330	1	AB026907	AB026907 Ochrobact
30	21	0.7	10353	1	AE001884	AE001884 Deinococc
31	21	0.7	38393	9	HSN62C4	Z73963 Human DNA s
32	21	0.7	68882	2	AC125428	AC125428 Homo sapi
33	21	0.7	110000	2	AL583837_1	Continuation (2 of
34	21	0.7	117951	9	AL359821	AL359821 Human DNA
35	21	0.7	126315	8	AF004274	AF004274 Oryza sat
36	21	0.7	133121	8	AC092553	AC092553 Oryza sat
37	21	0.7	134609	2	AC021416	AC021416 Homo sapi
38	21	0.7	135245	9	AL136092	AL136092 Human DNA
39	21	0.7	137724	8	AC122147	AC122147 Oryza sat
40	21	0.7	147207	2	AC078890	AC078890 Oryza sat
41	21	0.7	152120	9	AL391294	AL391294 Human DNA
42	21	0.7	182041	2	AC023025	AC023025 Homo sapi
43	21	0.7	182916	9	AC112482	AC112482 Homo sapi
44	21	0.7	196378	2	AC117504	AC117504 Homo sapi
45	21	0.7	208695	10	AL671478	AL671478 Mouse DNA

ALIGNMENTS

RESULT 1
HSA315733
Homo sapiens mRNA for metalloproteinase disintegrin 15 (ADAMTS15 gene).
2853 bp mRNA linear PRI 01-MAR-2002
LOCUS
DEFINITION
HSA315733
AJ315733.1 GI:19171175
VERSION
ADAMTS15 gene; disintegrin; metalloproteinase; thrombospondin.
KEYWORDS
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Cal,S., Obaya,A.J., Llamazares,M., Garabaya,C., Quesada,V. and Lopez-Otin,C.

TITLE Cloning, expression analysis, and structural characterization of seven novel human ADAMTS, a family of metalloproteinases with disintegrin and thrombospondin-1 domains

JOURNAL Gene 283 (1-2), 49-62 (2002)

MEDLINE 21856482

PUBMED 11867212

REFERENCE 2 (bases 1 to 2853)

AUTHORS Cal.S.

TITLE Direct Submission

JOURNAL Submitted (26-JUN-2001) Cal S., Biochemistry and Molecular Biology, University of Oviedo, Campus del Cristo, Asturias. 33006, SPAIN

FEATURES

source Location/Qualifiers

1..2853

/organism="Homo sapiens"

/db_xref="taxon:9606"

1..2853

/gene="ADAMTS15"

/gene="ADAMTS15"

/product="metalloprotease disintegrin 15 with thrombospondin domains"

/protein_id="CAC86014.1"

/db_xref="GI:19171176"

/translation="MLLLGLLLAFAGRTAGGSEPEREVVVPIRLDPDINGRRYYWRG PEDSGDGLIFQITAFQEDFYLHLPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCF YSDGVNAEPDSFAVSLCGGLRGAFGYRGAEYVISPLPNASAPAAQRNSQGAHLQRR YGPGSPGSDPSGVSAGWNPAIRLADPKPRAGFGESESRRSRAKRFVSIPIR YVTLVAADSESVKFGADLEHVLTLAARLXHPHSILNPINIVVKKVLLLRDRD SGPKVTGNALTLRNCAWQKLNKVDKHPYWDFAILFTQDLGCGATCTDGLMAD VTWCDFPSCVIEDGLPFAFTTAHELGHVFNPHDNVVKCEEVFGKLRANHMSP TLIQIDRANPWSASAIITDFLDSGGDLDDQPSKPSLPELDLPGLATLUSQOCEL AFVGSKPCPYMOYCTKLMTCKAKQOMVQTRHPFADGTSCGEGKLCGLKACVERH NLNKRVDGSWAKDPYGCSTCGGVQLARQCTNPTPANGRYCEGRVYKYSRN LEPSPSASGSPREBQCAFNGYNHSTNRLTLAVAWPKYSGVSPDRCKKILCRANG TGFEYVLAKVDGTLCSPDSTSVQVQKCIKAGCDNGLSKRRPKDCKVCGGDNKSC KKYTGFLTPMHGYNVVAIPAGASSIDIRORGYKGLIGDDNYLALKNSQKYLNLGH FVYSVERDLVWKSLLRSGTGATVESLQASRPILPELITVEVLSVGNMTPRRVYSF YLPKPERDKSVKDPGRPSVLNLSNQVEQPDPRFARVWAGSWGSPCSASCG SGLQKRVAGHSGAGQRTVACDAHRPVETQACGPEPTWELSAWSPCKSGRGFG RRLKVCVGHGRLRLDQNLHRKPOELDFCVLRPC"

BASE COUNT 522 a 921 c 903 g 507 t

ORIGIN

Query Match 96.2%; Score 2746; DB 9; Length 2853;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2796; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 57 TGAGCCAGAGCGGAGGTAGTCGTTCCATCCGACTCGACCGGACATTAACGGCGCGG 116

Db 57 TGAGCCAGAGCGGAGGTAGTCGTTCCATCCGACTCGACCGGACATTAACGGCGCGG 116

Qy 117 CTACTACTGCGGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCACAGC 176

Db 117 CTACTACTGCGGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCACAGC 176

Qy 177 ATTTACAGAGAGACTTTTACCTACACCTGACGCCGGATGCTCAGTTCTTGCGTCCCGCCT 236

Db 177 ATTTACAGAGAGACTTTTACCTACACCTGACGCCGGATGCTCAGTTCTTGCGTCCCGCCT 236

Qy 237 CTCACACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGGG 296

Db 237 CTCACACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGGG 296

Qy 297 ACGTGTCTTATCTCGGGACGTGAACCGCGAGCGGAGCTCGTTCGCTGCTGTGAGCCT 356

Db 297 ACGTGTCTTATCTCGGGACGTGAACCGCGAGCGGAGCTCGTTCGCTGCTGTGAGCCT 356

Qy 357 GTGGGGGGGCTCCCGGAGCCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCGCT 416

Db 357 GTGGGGGGGCTCCCGGAGCCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCGCT 416

Qy 417 GCCCAATGCTAGCGCGCGCGGCGGCGCAACAGCCAGCGGCGCACACCTTCTCCAGCG 476

Db 417 GCCCAATGCTAGCGCGCGCGGCGGCGCAACAGCCAGGCGGCACACCTTCTCCAGCG 476

Qy 477 CCGGGGTGTTCCGGGGCGGGCCCTTCGCGAGAGCCACACCTTCTCGCTCGCGGGTGGGCTCGGG 536

Db 477 CCGGGGTGTTCCGGGGCGGGCCCTTCGCGAGAGCCACACCTTCTCGCTCGCGGGTGGGCTCGGG 536

Qy 537 CTGGAACCCCGCCATCTTACGGGCGCTTGAGACCCCTTACAAAGCGGCGGGGCGGGCTTCGG 596

Db 537 CTGGAACCCCGCCATCTTACGGGCGCTTGAGACCCCTTACAAAGCGGCGGGGCGGGCTTCGG 596

Qy 597 GGAGAGTGTAGTCCGGCGGAGGTCTGGGGCGGCGCAAGCGTTTCGTGTCTATCCCGCGGTA 656

Db 597 GGAGAGTGTAGTCCGGCGGAGGTCTGGGGCGGCGCAAGCGTTTCGTGTCTATCCCGCGGTA 656

Qy 657 CGTGAGAGCGTGTGTGTCGCGGAGAGTCAATGTGTCAAAGTTTCCACGCGCGGGAGCTTGA 716

Db 657 CGTGAGAGCGTGTGTGTCGCGGAGAGTCAATGTGTCAAAGTTTCCACGCGCGGGAGCTTGA 716

Qy 717 ACATATATCTGTCGCTGTGGCAACCGGCGGCGGAGCTTACCGCCATCCAGCATCT 776

Db 717 ACATATATCTGTCGCTGTGGCAACCGGCGGCGGAGCTTACCGCCATCCAGCATCT 776

Qy 777 CAACCCCATCAACATCGTTGTGTCAAGTGCTGCTTCTTAGAGATCTGACTCCGGGCC 836

Db 777 CAACCCCATCAACATCGTTGTGTCAAGTGCTGCTTCTTAGAGATCTGACTCCGGGCC 836

Qy 837 CAAGGTACCGGCAATGCGGCGCTGACGCTGCGCAACTTCTGTGCTGGCAGAAAGCT 896

Db 837 CAAGGTACCGGCAATGCGGCGCTGACGCTGCGCAACTTCTGTGCTGGCAGAAAGCT 896

Qy 897 GAACAAAGTGTAGTCAACAGCACCCCGAGTACTGGGACACTGCCATCTTCTTACACAGCA 956

Db 897 GAACAAAGTGTAGTCAACAGCACCCCGAGTACTGGGACACTGCCATCTTCTTACACAGCA 956

Qy 957 GGACCTGTGTGAGCGACACCACTGTGACACCTGGGCACTGGCTGATGTGGGTACCATGTG 1016

Db 957 GGACCTGTGTGAGCGACACCACTGTGACACCTGGGCACTGGCTGATGTGGGTACCATGTG 1016

Qy 1017 TGACCCCAAGAGAAGTGTCTGTCTATTGAGGACGATGGGCTTCCATCAGCCTTCCACCAC 1076

Db 1017 TGACCCCAAGAGAAGTGTCTGTCTATTGAGGACGATGGGCTTCCATCAGCCTTCCACCAC 1076

Qy 1077 TGCCACAGAGTGTGGCGCGCTGTCAACATGCCCATGCAATGTGAAGTCTGTGAGGA 1136

Db 1077 TGCCACAGAGTGTGGCGCGCTGTCAACATGCCCATGCAATGTGAAGTCTGTGAGGA 1136

Qy 1137 GGTGTTTGGGAAGCTCCGAGGCAACACATGATGTCCCGACCCCTCATCCAGATCGACCG 1196

Db 1137 GGTGTTTGGGAAGCTCCGAGGCAACACATGATGTCCCGACCCCTCATCCAGATCGACCG 1196

Qy 1197 TGCCAAACCCCTGGTCAGCCTGCGAGTGTGCCATCATCACCAGCTTCTTGGACAGCGGCA 1256

Db 1197 TGCCAAACCCCTGGTCAGCCTGCGAGTGTGCCATCATCACCAGCTTCTTGGACAGCGGCA 1256

Qy 1257 CGGTGACTGCTTCCGCAACCAAGCCATCTCCCTGCGCCGAGGATCTCCCGGG 1316

Db 1257 CGGTGACTGCTTCCGCAACCAAGCCATCTCCCTGCGCCGAGGATCTCCCGGG 1316

Qy 1317 CGCAGCTACACCCCTGAGCAGCAGTGCAGCTGGCTTTTGGCGTGGGCTCCAAAGCCCTG 1376

Db 1317 CGCAGCTACACCCCTGAGCAGCAGTGCAGCTGGCTTTTGGCGTGGGCTCCAAAGCCCTG 1376

Qy 1377 TCGTTACATGCACTACTGACCAAGCTGTGGTGCACCGGGAAGGCAAGGACAGATGT 1436

Db 1377 TCGTTACATGCACTACTGACCAAGCTGTGGTGCACCGGGAAGGCAAGGACAGATGT 1436

Qy 1437 GTGCCAGACCCGCACTTCCCTCGTGGCCGATGGCACAGCTGTGGCGAGGCAAGCTCTG 1496

Db 1437 GTGCCAGACCCGCACTTCCCTCGTGGCCGATGGCACAGCTGTGGCGAGGCAAGCTCTG 1496

Qy 1497 CCTCAAAGGGGCTCGGTGGAGAGACAACTCTCAAAAGCACAGGCTGATGTTCTCTG 1556

Db 1497 CCTAAAGGGGCGCTGCTGGAGAGACACAACTCAACAAGCACAGGCTGGATGTTCTCTG 1556
QY 1557 GSCAAATGGGATCCCTATGCGCCCTGCTCGGCACATGTTGGGGCGGTGCAAGCTGGC 1616
Db 1557 GSCAAATGGGATCCCTATGCGCCCTGCTCGGCACATGTTGGGGCGGTGCAAGCTGGC 1616
QY 1617 CAGGAGGAGTGCACAAACCCACCCCTGCCAAAGGGGGCAAGTACTCGAGGGAGTGA 1676
Db 1617 CAGGAGGAGTGCACAAACCCACCCCTGCCAAAGGGGGCAAGTACTCGAGGGAGTGA 1676
QY 1677 GGTAAATACGATGCTCAACCTGGAGCCCTGCCAGCTCAGCTCCGGAAGAGCTT 1736
Db 1677 GGTAAATACGATGCTCAACCTGGAGCCCTGCCAGCTCAGCTCCGGAAGAGCTT 1736
QY 1737 CGGAGGAGAGTGTGAGGCTTCAACGGCTACAAAGGCTACAAAGGCTACAGCTC 1796
Db 1737 CGGAGGAGAGTGTGAGGCTTCAACGGCTACAAAGGCTACAAAGGCTACAGCTC 1796
QY 1797 CGCGTGGCATGGGTGCCAAGTACTCGGGGCTGTCTCCCGGGACAAAGTGCAGAGCTCAT 1856
Db 1797 CGCGTGGCATGGGTGCCAAGTACTCGGGGCTGTCTCCCGGGACAAAGTGCAGAGCTCAT 1856
QY 1857 CTGCGAGCCAAATGCACTGCTTCTATGCTGTCACCAAGGCTGTCAGAGGCTGTCAG 1916
Db 1857 CTGCGAGCCAAATGCACTGCTTCTATGCTGTCACCAAGGCTGTCAGAGGCTGTCAG 1916
QY 1917 GCTGTGCTCTCTGACTCCACTCGCTGCTGTCACCAAGGCTGTCAGAGGCTGTCAG 1976
Db 1917 GCTGTGCTCTCTGACTCCACTCGCTGCTGTCACCAAGGCTGTCAGAGGCTGTCAG 1976
QY 1977 TGATGGGAACCTGGGCTCCAAAGAGATTCGACAAGTGTGGGTGTGTGGGGGAGACAA 2036
Db 1977 TGATGGGAACCTGGGCTCCAAAGAGATTCGACAAGTGTGGGTGTGTGGGGGAGACAA 2036
QY 2037 TAAGAGCTGCAAGAGGCTGCTGCTTCTTCAACCAAGGCTGTCAGAGGCTGTCAGAG 2096
Db 2037 TAAGAGCTGCAAGAGGCTGCTGCTTCTTCAACCAAGGCTGTCAGAGGCTGTCAGAG 2096
QY 2097 GGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCCCGCAGCGGCTTCAAGAGGCT 2156
Db 2097 GGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCCCGCAGCGGCTTCAAGAGGCT 2156
QY 2157 GATCGGGATGACAACTACCTGGCTCTGAAGAAGCAAGCAAGTACCTGCTCAAGCG 2216
Db 2157 GATCGGGATGACAACTACCTGGCTCTGAAGAAGCAAGCAAGTACCTGCTCAAGCG 2216
QY 2217 GCATTTGCTGCTGCTGGGCTGGAGCGGCTGCTGCTGAGGCGACTGCTGCTGCGGTA 2276
Db 2217 GCATTTGCTGCTGCTGGGCTGGAGCGGCTGCTGCTGAGGCGACTGCTGCTGCGGTA 2276
QY 2277 CAGCGCACGGGCACAGCGGCTGGAGCGCTGAGGCTTCCCGGCGCCATCCTGGAGCGCT 2336
Db 2277 CAGCGCACGGGCACAGCGGCTGGAGCGCTGAGGCTTCCCGGCGCCATCCTGGAGCGCT 2336
QY 2337 GACGCTGAGGCTCTCTCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTTCTTA 2396
Db 2337 GACGCTGAGGCTCTCTCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTTCTTA 2396
QY 2397 TCTGCCAAGAGCTCGGGAGGACAAGTCTCTCATCCCAAGGACCCCGGGGACCTC 2456
Db 2397 TCTGCCAAGAGCTCGGGAGGACAAGTCTCTCATCCCAAGGACCCCGGGGACCTC 2456
QY 2457 TGTCTTGCAACAGCGCTCTCAGCTCTTCCAAAGGCTGGAGCAGCGGACAGAGGCC 2516
Db 2457 TGTCTTGCAACAGCGCTCTCAGCTCTTCCAAAGGCTGGAGCAGCGGACAGAGGCC 2516
QY 2517 CCTGCAAGCTGGGTGGCTGAGCTGGGGCGGCTGCTCCGCGAGCTCGGCAAGTGGCT 2576
Db 2517 CCTGCAAGCTGGGTGGCTGAGCTGGGGCGGCTGCTCCGCGAGCTCGGCAAGTGGCT 2576
QY 2577 GCAGAGCGGGGCTGAGCTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCGCTGGA 2636
Db 2577 GCAGAGCGGGGCTGAGCTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCGCTGGA 2636

QY 2637 TGCAGCCCATGCGCCCTGGAGACAAAGCTTCCGGGAGCCCTGCCCACTGCGAGCT 2696
Db 2637 TGCAGCCCATGCGCCCTGGAGACAAAGCTTCCGGGAGCCCTGCCCACTGCGAGCT 2696
QY 2697 CAGCGCTGCTCACCCTGCTCCAAAGCTGCGGCCGGGATTTACAGAGCGCTCACTAA 2756
Db 2697 CAGCGCTGCTCACCCTGCTCCAAAGCTGCGGCCGGGATTTACAGAGCGCTCACTAA 2756
QY 2757 GTGTGTGGCCACGAGCGCGCTGCTGCGCGGACCACTGCAACTTGCACCGCAAGCC 2816
Db 2757 GTGTGTGGCCACGAGCGCGCTGCTGCGCGGACCACTGCAACTTGCACCGCAAGCC 2816
QY 2817 CCAGGAGCTGACTTCTGCTGCTGCTGAGGCGCTGCTGA 2853
Db 2817 CCAGGAGCTGACTTCTGCTGCTGCTGAGGCGCTGCTGA 2853
RESULT 2
AX342635 2930 bp DNA Linear PAT 12-JAN-2002
LOCUS AX342635 Sequence 32 from Patent WO0198468.
DEFINITION AX342635
ACCESSION AX342635
VERSION AX342635.1 GI:18152032
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Yue, H., Elliott, V.S., Gandhi, A.R., Lal, P., Au-Young, J.,
Tribouley, C.M., Deleage, A.M., Baughn, M.R., Nguyen, D.B., Lee, E.A.,
Hafalia, A., Khan, F.A., Walla, N.K., Yao, M.G., Lu, D.A., Patterson, C.,
Tang, Y.T., Walsh, R.T., Azimzai, Y., Rammakumar, J., Xu, Y. and Reddy, R.
JOURNAL Patent: WO 0198468-A 32 27-DEC-2001;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
source
1..2930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 7473089CB1"
BASE COUNT 529 a 958 c 931 g 512 t
ORIGIN
Query Match 73.2%; Score 2089; DB 6; Length 2930;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2379; Conservative 0; Mismatches 1; Indels 3; Gaps 1;
QY 57 TGAGCCAGAGCGGAGTACTGTTCCATCCGACTGCGGCCGACATTAACGGCGCGC 116
Db 131 TGAGCCAGAGCGGAGTACTGTTCCATCCGACTGCGGCCGACATTAACGGCGCGC 190
QY 117 CTACTACTGCGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTCAGATACAGC 176
Db 191 CTACTACTGCGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTCAGATACAGC 250
QY 177 ATTTTCAGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCCGCCT 236
Db 251 ATTTTCAGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCCGCCT 310
QY 237 CTCACCTGAGCATCTGGGCTGCCCTCCAGGGGCTCACCGGGGCTTTCAGACCTGCG 296
Db 311 CTCACCTGAGCATCTGGGCTGCCCTCCAGGGGCTCACCGGGGCTTTCAGACCTGCG 370
QY 297 ACCTGCTTCTATTTCTGGGACCTGAACCGCGGACTCGTTCGCTCTGCTGAGCCT 356
Db 371 ACCTGCTTCTATTTCTGGGACCTGAACCGCGGACTCGTTCGCTCTGCTGAGCCT 430
QY 357 GTGCGGGGCTCCCGCGAGCTTTTGGCTACCGAGGCGCGAGTATGCTCTTAGTCGCT 416
Db 431 GTGCGGGGCTCCCGCGAGCTTTTGGCTACCGAGGCGCGAGTATGCTCTTAGTCGCT 490
QY 417 GCCCAATGCTAGCGCGCGGCGCGACAGCCAGCGGCGCACACCTTCTCCAGCG 476

Db 491 G C C C A A T G T A G C G C G C G G C G C A G C G C A A C A G C A G G C G C A C A C C T T C T C C A G C G 550
Qy 477 C G G G G T T T C C G G G C G G G C C T T C C G G A G A C C C C A C C T C T C G T C G C G G G T G G C C T C G G G 536
Db 551 C G G G G T T C C G G G C G G C C T T C C G G A G A C C C C A C C T C T C G T C G C G G G T G G C C T C G G G 610
Qy 537 C T G A A C C C C C C A T C T A C G G G C C C T G G A C C C T T A A G C C G G G G C G G C G G C T T C G G 596
Db 611 C T G A A C C C C C C A T C T A C G G G C C C T G G A C C C T T A A G C C G C G G G C G G C T T C G G 670
Qy 597 G G A G A G T G T A G C G G C G G C G G C T G G C G G C C A G C G T T C G T G C T A T C C G C G G T A 656
Db 671 G G A G A G T G T A G C G G C G G C G G C T G G C G G C C C A G C G T T C G T G C T A T C C G C G G T A 730
Qy 657 C G T G G A G A C C T G T G T G C G G A G C A G T C A A T G G T C A A G T T T C A C G C G C G C G A C C T G G A 716
Db 731 C G T G G A G A C C T G T G T G C G G A G C A G T C A A T G G T C A A G T T T C A C G C G C G C G A C C T G G A 790
Qy 717 A C A T T A T C T G T A G C T G C T G G C A A C G G C G G C G A C T C T A C G C C A T C C C A G C A T C C T 776
Db 791 A C A T T A T C T G T A G C T G T G G C A A C G G C G G C G A C T C T A C G C C A T C C C A G C A T C C T 850
Qy 777 C A A C C C C A C A A C A T C G T T G T G T C A A G T G C T G C T T C T T A G A G A T C G T A C T C C G G C C 836
Db 851 C A A C C C C A C A A C A T C G T T G T G T C A A G T G C T G C T T C T T A G A G A T C G T A C T C C G G C C 910
Qy 837 C A A G T C A C C G G C A A T C G G C C C T G A C C T G C G C A A C T T C T G T G C C T G G C A G A A G C T 896
Db 911 C A A G T C A C C G G C A A T C G G C C C T G A C C T G C G C A A C T T C T G T G C C T G G C A G A A G C T 970
Qy 897 G A A C A A A G T A G T G A C A A G C A C C C G A G T A C T G G G A C A C T G C C A T C C T T C A C A G G C A 956
Db 971 G A A C A A A G T A G T G A C A A G C A C C C G A G T A C T G G G A C A C T G C C A T C C T C T T C A C A G G C A 1030
Qy 957 G G A C C T G T G T G G A C C A C C A C C T G T G A C A C C C T G G G A T G C T G A T G T G G G T A C C A T G T G 1016
Db 1031 G G A C C T G T G T G G A C C A C C A C C T G T G A C A C C C T G G G A T G C T G A T G T G G G T A C C A T G T G 1090
Qy 1017 T G A C C C C A A G A A G A C T G C T G T C A T T A G A G A C G A T G G G C T T C C A T A G C C T T C A C C A C 1076
Db 1091 T G A C C C C A A G A A G A C T G C T G T C A T T A G A G A C G A T G G G C T T C C A T A G C C T T C A C C A C 1150
Qy 1077 T G C C A C A G A C T G G C C C A C C G T T T C A A C A T G C C C C A T G A C A A T G T A A A G T C T G T C A G G A 1136
Db 1151 T G C C C A C A G A C T G G C C A C C G T T T C A C A T G C C C C A T G A C A A T G T A A A G T C T G T C A G G A 1210
Qy 1137 G G T G T T T G G A A G T C C G A G C C A A C C A C A T G A T G T C C C G A C C C T C A T C C A G A T C G A C C G 1196
Db 1211 G G T G T T T G G A A G T C C G A G C C A A C C A C A T G A T G T C C C G A C C C T C A T C C A G A T C G A C C G 1270
Qy 1197 T G C C A A C C C T G T A G C C T G C A G T G T G C C A T C A T C A C G G A C T T C C T G G A C A G C G G C A 1256
Db 1271 T G C C A A C C C T G T G A C C T G C A G T G C T G C A T C A T C A C G G A C T T C C T G G A C A G C G G C A 1330
Qy 1257 C G G T A C T G C C T C T G A C C A A C C C A G A C C C A C T C C C T G C C C G A G A T C T G C C G G G 1316
Db 1331 C G G T A C T G C C T C T G A C C A A C C C A G A C C C A C T C C C T G C C C G A G A T C T G C C G G G 1390
Qy 1317 C G C A G A C C C C A C C T T C C C C T G G G C C G A T G G C A C C A G C T G T G G C G G G C T C C A A G C C C T G 1376
Db 1391 C G C A G A C C C C A C C T T C C C C T G G C C G A T G G C A C C A G C T G T G G C G A G G C C A G C T C T G 1450
Qy 1377 T C C T T A C A T G A C T A C C A C C A G C T G T G T G C A C C G G G A A G C C C A A G G A C A G A T G T 1436
Db 1451 T C C T T A C A T G A C T A C C A C C A A G C T G T G T G C A C C G G G A A G C C C A A G G A C A G A T G T 1510
Qy 1437 G T G C A G A C C C G C A C T T C C C C T G G G C C G A T G G C A C C A G C T G T G G C G A G G C A A G C T C T G 1496
Db 1511 G T G C A G A C C C G C A C T T C C C C T G G C C G A T G G C A C C A G C T G T G G C G A G G C C A G C T C T G 1570
Qy 1497 C C T C A A A G G G C C T G C G T G G A G A C A C A C C A C C A A G C A C A G G T G G A T G G T T C C T G 1556

Db 1571 C C T C A A A G G G C C T G C G T G G A G A G A C A A C C T C A A A G A C A C A G G T G G A T G G T T C C T G 1630
Qy 1557 G G C C A A T G G A T C C C T A T G G C C C T G C T G C G C A C A T G T G T G G G G C G C T G C A C T G C G 1616
Db 1631 G G C C A A T G G A T C C C T A T G G C C C T G C T G C G C A C A T G T G T G G G G C G C T G C A C T G C G 1690
Qy 1617 C A G G A G C A G T C A C C A A C C C C C A C C C C T G C C A A C G G G G C A A G T A C T C G A G G G A G T G A G 1676
Db 1691 C A G G A G C A G T C A C C A A C C C C C A C C C C T G C C A A C G G G G C A A G T A C T C G A G G A G T G A G 1750
Qy 1677 G G T G A A F A C C A T C C T G C A A C C T G G A G C C T G C C C A G C T C A G C C T C C G N A A G A G C T T 1736
Db 1751 G G T G A A T A C C A T C C T G C A A T C T G A G C C T G C C C A G C T C A G C C T C C G A A A G A G C T T 1810
Qy 1737 C C G G A G A G A C A G T G T A G C C T T T C A A C G G C T A C A C C A C A C A G C A C C A C C G G C T C A C T C T 1796
Db 1811 C C G G A G A G A C A G T G T A G C C T T T C A A C G G C T A C A C C A C A C A G C A C C A C C G C T C A C T C T 1870
Qy 1797 C C G C T G G C A T G G T G C C C A A G T A C T C C G G C T G T C T C C C G G G A C A A G T G C A A G C T C A T 1856
Db 1871 C G C C T G G C A T G G T G C C C A A G T A C T C C G G C T G T C T C C C G G G A C A A G T G C A A G C T C A T 1930
Qy 1857 C T G C C G A G C A A T G C A C T G G C T A C T T C T A T G T G C T G C A C C C A A - - - G G T G G T G A C G G 1913
Db 1931 C T G C C A G C A A T G G C A T G G C T A C T T C T A T G T G C T G G C A C C A A G G T G G T G G A C G G 1990
Qy 1914 C A C G T G T C T C T C C T G A C T C C A C C T C G T C T G T T C C A A G C C A A G T G C A T C A A G C T G G 1973
Db 1991 C A C G T G T C T C T C T G A C T C C A C C T C G T C T G T T C C A A G C C A A G T G C A T C A A G C T G G 2050
Qy 1974 C T G T A T G G G A C C T G G G C T C C A A A G A G A T T C A C A A G T G T G G G T G T G T G G G G A G A 2033
Db 2051 C T G T A T G G G A C C T G G G C T C C A A A G A G A T T C G A C A A G T G T G G G T G T G T G G G G A G A 2110
Qy 2034 C A A T A A G A C T C A A A A A G G T G A C T G G A C T C T T C A C A A G C C C A T G C A T G C T A C A A T T T 2093
Db 2111 C A A T A A G A C T C A A A A A G G T G A C T G G A C T C T T C A C A A G C C C A T G C A T G C T A C A A T T T 2170
Qy 2094 C G T G T G G C C A T C C C G A G C G C C T C A A G C A T C G A C A T C C C A G C C G C G T T A C A A A G G 2153
Db 2171 C G T G T G G C C A T C C C G A G C G C C T C A A G C A T C G A C A T C C C C A G C G C G T T A C A A A G G 2230
Qy 2154 G C T G A T C G G G A T A C A A C T A C T G G C T C T G A A G A C A C C A A G C A A G C A A G T A C C T G C T C A A 2213
Db 2231 G C T G A T C G G G A T A C A A C T A C T G G C T C T G A A A A C A G C C A A G C A A G T A C C T G C T C A A 2290
Qy 2214 C G G G C A T T T C G T G G T G C G C G G T G G A C C G G A C C T G G T G T G A A G G C A G T C T G C T G C G 2273
Db 2291 C G G G C A T T T C G T G G T G C G C G G T G G A G C G G A C C T G G T G T G A A G G C A G T C T G C T G C G 2350
Qy 2274 G T A C A G C G C A G C G G C A C A G C G G T G G A G A C C T G C A G C C T T C C C G G C C C A T C C T G G A G C C 2333
Db 2351 G T A C A G C G C A G C G G C A C A G C G G T G G A G A C C T G C A G C C T T C C C G G C C C A T C C T G G A G C C 2410
Qy 2334 G C T G A C C T G G A G G T C C T C C G T G G G A A G A T G A C A C C C G C G G G T C C C G C T A C T C C T T 2393
Db 2411 G C T G A C C T G G A G G T C C T C C G T G G G A A G A T G A C A C C C C C C G G G T C C C G C T A C T C C T T 2470
Qy 2394 C T A T C T C C C A A A G A C C T C G G A G A C A A G T C C T C T C A T C C C 2436
Db 2471 C T A T C T C C C A A A G A C C T C G G A G A C A A G T C C T C T C A T C C C 2513

RESULT 3
AX319860

LOCUS

DEFINITION Sequence 24 from Patent WO0183782.

ACCESSION AX319860

VERSION AX319860.1 GI:17901450

KEYWORDS

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1
AUTHORS Plozman,G.D., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and Payne,V.
TITLE Novel proteases
JOURNAL Patent: WO 0183782-A 24 08-NOV-2001;
Sugen, Inc. (US)
FEATURES Location/Qualifiers
source 1..2937
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 531 a 958 c 936 g 512 t
ORIGIN

Query Match 73.2%; Score 2089; DB 6; Length 2937;
Best Local Similarity 99.8%; Pred No. 0;
Matches 2379; Conservative 0; Mismatches 1; Indels 3; Gaps 1;
QY 57 TGAGCCAGAGCGGAGGTAGTGGTCCCATCCGACTGGACCCGGACATTAAACGGCGCGG 116
DB 132 TGAGCCAGAGCGGAGGTAGTGGTCCCATCCGACTGGACCCGGACATTAAACGGCGCGG 191
QY 117 CTACTACTGCGGGGTCCGAGAGACTCCGGGATCAGGACTCATTTTCAGATCACAGC 176
DB 192 CTACTACTGCGGGGTCCGAGAGACTCCGGGATCAGGAGCTCATTTTCAGATCACAGC 251
QY 177 ATTTCAGGAGGACTTTTACCTACACCTAGCGCGGATGCTCAGTTCTTGGCTCCCGCCTT 236
DB 252 ATTTCAGGAGGACTTTTACCTACACCTAGCGCGGATGCTCAGTTCTTGGCTCCCGCCTT 311
QY 237 CTCACCTGAGCATCTGGGGTCCCTCCAGGGGTACCGGGGCTCTTCAGACCTGCG 296
DB 312 CTCACCTGAGCATCTGGGGTCCCTCCAGGGGTACCGGGGCTCTTCAGACCTGCG 371
QY 297 ACCTGCTTCTATTCTGGGAGGTGAACGCCAGCGCGGACTGCTCGCTGCTGTAGCCT 356
DB 372 ACCTGCTTCTATTCTGGGAGGTGAACGCCAGCGCGGACTGCTCGCTGCTGTAGCCT 431
QY 357 GTGCGGGGGGCTCCGCGGAGCCTTTGGCTACCGAGGCGCGAGTAGTATGATTTAGCCCGCT 416
DB 432 GTGCGGGGGGCTCCGCGGAGCCTTTGGCTACCGAGGCGCGAGTAGTATGATTTAGCCCGCT 491
QY 417 GCCCAATGCTAGCGCGCGCGCGCAGCGGCAACAGCCAGCGGCGACACCTTCTCCAGG 476
DB 492 GCCCAATGCTAGCGCGCGCGCGCAGCGGCAACAGCCAGCGGCGACACCTTCTCCAGG 551
QY 477 CCGGGGTGTTCGGGGGGGCTTCGGGAGACCCACCTCTCGCTCGGGGGTGGCTCGGG 536
DB 552 CCGGGGTGTTCGGGGGGGCTTCGGGAGACCCACCTCTCGCTCGGGGGTGGCTCGGG 611
QY 537 CTGGAACCCCGCATCTTACGGGCGCTTACAGCCCTTACAGCGCGCGGGGGGCTTCGG 596
DB 612 CTGGAACCCCGCATCTTACGGGCGCTTACAGCCCTTACAGCGCGCGGGGGGCTTCGG 671
QY 597 GGAGAGCTGTAGCGCGCGAGGTCTGGGGCGCGCAAGGCTTTCGCTATCCCGCGGTA 656
DB 672 GGAGAGCTGTAGCGCGCGAGGTCTGGGGCGCGCAAGGCTTTCGCTATCCCGCGGTA 731
QY 657 CBTGGAGACGCTGGTGGTCCGAGAGGTCAATGGTTCAGTTCCAGCGCGCGGACCTTGA 716
DB 732 CBTGGAGACGCTGGTGGTCCGAGAGGTCAATGGTTCAGTTCCAGCGCGCGGACCTTGA 791
QY 717 ACATTATCTGCTGACGCTGCTGGCAACGGCGCGGACTCTTACCGCCATCCAGCATCCT 776
DB 792 ACATTATCTGCTGACGCTGCTGGCAACGGCGCGGACTCTTACCGCCATCCAGCATCCT 851
QY 777 CAACCCCATCAACATCTGTTGGTCAAGGTGCTGCTTACAGATCTGACTCTCGGGCC 836
DB 852 CAACCCCATCAACATCTGTTGGTCAAGGTGCTGCTTACAGATCTGACTCTCGGGCC 911
QY 837 CAAGGTACCGGCAATGCGGGCCCTGACGCTGCGCAACTTCTGTGCTTGGCAAGAAAGCT 896
DB 912 CAAGGTACCGGCAATGCGGGCCCTGACGCTGCGCAACTTCTGTGCTTGGCAAGAAAGCT 971

QY 897 GAACAAAGTGTAGTGACAAAGCACCCCGAGTACTGGGACACTGCCATCTCTTTCACAGGCA 956
DB 972 GAACAAAGTGTAGTGACAAAGCACCCCGAGTACTGGGACACTGCCATCTCTTTCACAGGCA 1031
QY 957 GGACCTGTGTGGAGCACCACTGTGTACACCTCGGGATGCTGTGGTGGTACCATGTG 1016
DB 1032 GGACCTGTGTGGAGCACCACTGTGTACACCTCGGGATGCTGTGGTGGTACCATGTG 1091
QY 1017 TGACCCCAAGAGAAAGCTGCTGTCTGATTTAGGAGCATGGGCTTCATCAGCTTTCACAC 1076
DB 1092 TGACCCCAAGAGAAAGCTGCTGTCTGATTTAGGAGCATGGGCTTCATCAGCTTTCACAC 1151
QY 1077 TGCCACAGAGCTGGGCGGACGCTGTTCAACATGCCCATGACAAATGTAAAGTCTGTGAGGA 1136
DB 1152 TGCCACAGAGCTGGGCGGACGCTGTTCAACATGCCCATGACAAATGTAAAGTCTGTGAGGA 1211
QY 1137 GGTGTTTGGGAAGCTCCGAGCAACACCATGATGTCCCGGACCTCATCAGATCGACCG 1196
DB 1212 GGTGTTTGGGAAGCTCCGAGCAACACCATGATGTCCCGGACCTCATCAGATCGACCG 1271
QY 1197 TGCCAAACCCCTGTGCTGAGCTGCTGCTGATCATCAGGACTTCTTGACAGCGGGCA 1256
DB 1272 TGCCAAACCCCTGTGCTGAGCTGCTGCTGATCATCAGGACTTCTTGACAGCGGGCA 1331
QY 1257 CGGTGACTGCTCTCTGGACCAACCCAGCAAGCCCATCTCCCTGCGCGGAGATCTGCCGG 1316
DB 1332 CGGTGACTGCTCTCTGGACCAACCCAGCAAGCCCATCTCCCTGCGCGGAGATCTGCCGG 1391
QY 1317 CGCAGCTTACACCTGAGCCAGCAGTGCAGCTGGCTTTTGGCTGGGCTCCAGGCCCTG 1376
DB 1392 CGCAGCTTACACCTGAGCCAGCAGTGCAGCTGGCTTTTGGCTGGGCTCCAGGCCCTG 1451
QY 1377 TCCTTACATGCACTACTGCAACCAAGCTGTGGTGACCGGGAAGCCAAAGGACAGATGCT 1436
DB 1452 TCCTTACATGCACTACTGCAACCAAGCTGTGGTGACCGGGAAGCCAAAGGACAGATGCT 1511
QY 1437 GTGCCAGACCCCGCCTTCCCTTGGCGGATGGCAGCAGCTGTGGCGAGGCGCAAGCTCTG 1496
DB 1512 GTGCCAGACCCCGCCTTCCCTTGGCGGATGGCAGCAGCTGTGGCGAGGCGCAAGCTCTG 1571
QY 1497 CCTCAAAGGGGCTGTGGGAGACACACCTCAACAGCACAGGCTGATGCTGCTG 1556
DB 1572 CCTCAAAGGGGCTGTGGGAGACACACCTCAACAGCACAGGCTGATGCTGCTG 1631
QY 1557 GCCCAATGGGATCCCTTATGGCCCTTCTCGCGCACATGTGGTGGGGCGTGCAGCTGGC 1616
DB 1632 GCCCAATGGGATCCCTTATGGCCCTTCTCGCGCACATGTGGTGGGGCGTGCAGCTGGC 1691
QY 1617 CAGGAGCAGTGCACCAACCCCTGCGCAACGGGGCAAGTACTGCGAGGAGTGA 1676
DB 1692 CAGGAGCAGTGCACCAACCCCTGCGCAACGGGGCAAGTACTGCGAGGAGTGA 1751
QY 1677 GGTGAATACCGATCCCTGCAACCTGGAGCCCTGCCCGAGCTCAGCTTCCCGAAGAGCTT 1736
DB 1752 GGTGAATACCGATCCCTGCAACCTGGAGCCCTGCCCGAGCTCAGCTTCCCGAAGAGCTT 1811
QY 1737 CCGGAGGAGCAGTGTGAGGCTTTCACAGGCTTCAACACACAGCACCAACCGGCTCACTCT 1796
DB 1812 CCGGAGGAGCAGTGTGAGGCTTTCACAGGCTTCAACACACAGCACCAACCGGCTCACTCT 1871
QY 1797 CGCGGTGGCATGGGTGCCCAAGTACTCCGGCGTGTCTCCCGGGGACAAGTGCAGGCTCAT 1856
DB 1872 CGCGGTGGCATGGGTGCCCAAGTACTCCGGCGTGTCTCCCGGGGACAAGTGCAGGCTCAT 1931
QY 1857 CTGCCGAGCCAAATGGCACTGGCTACTTCTATGTGCTGGCACCCAA ---GGTGTGACGG 1913
DB 1932 CTGCCGAGCCAAATGGCACTGGCTACTTCTATGTGCTGGCACCCAAAGGCTGGTGGTGGACGG 1991
QY 1914 CACGCTGTGCTCTCTGCTGACTCCACTCCGCTGTGTGTCGAAGGAGTGCATCAAGCTGG 1973
DB 1992 CACGCTGTGCTCTCTGCTGACTCCACTCCGCTGTGTGTCGAAGGAGTGCATCAAGGCTGG 2051

```
QY 1974 CTGTGATGGGAACCTGGGCTCCAGAGAGATTCGACAAAGTGTGGGTGTGGGGAGA 2033
Db 2052 CTGTGATGGGAACCTGGGCTCCAGAGAGATTCGACAAAGTGTGGGTGTGGGGAGA 2111
QY 2034 CAATAAGAGCTGCAAGAGGTGACTGGACTCTTCACCAAGCCCATGCATGGCTACAAATTT 2093
Db 2112 CAATAAGAGCTGCAAGAGGTGACTGGACTCTTCACCAAGCCCATGCATGGCTACAAATTT 2171
QY 2094 CGTGGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCCGCGACGCGGTTTACAAAGG 2153
Db 2172 CGTGGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCCGCGACGCGGTTTACAAAGG 2231
QY 2154 GCTGATCGGGGATGACAACCTACCTGGCTCTGGAAGAACAGCCAAAGCAAGTACCTGCTCAA 2213
Db 2232 GCTGATCGGGGATGACAACCTACCTGGCTCTGGAAGAACAGCCAAAGCAAGTACCTGCTCAA 2291
QY 2214 CGGCGATTTCTGGTGTGGCGGTGAGCGGGACCTGTGGTGAAGGGAGTCTGCTGG 2273
Db 2292 CGGCGATTTCTGGTGTGGCGGTGAGCGGGACCTGTGGTGAAGGGAGTCTGCTGG 2351
QY 2274 GTACAGCGGCACGGGCACAGCGGTGAGAGCCTGCAAGCTTCCCGGCCCATCTCGTAGCG 2333
Db 2352 GTACAGCGGCACGGGCACAGCGGTGAGAGCCTGCAAGCTTCCCGGCCCATCTCGTAGCG 2411
QY 2334 GCTGACCGTGGAGTCTCTCTCGTGGGGAAGATGACACCGCCCGCGGTCCGCTACTCCTT 2393
Db 2412 GCTGACCGTGGAGTCTCTCTCGTGGGGAAGATGACACCGCCCGCGGTCCGCTACTCCTT 2471
QY 2394 CTATCTGCCCAAGAGCCTCGGAGGACCAAGTCTCTCATCCC 2436
Db 2472 CTATCTGCCCAAGAGCCTCGGAGGACCAAGTCTCTCATCCC 2514
```

```
RESULT 4
AP002986 157963 bp DNA linear PRI 08-MAR-2002
LOCUS Homo sapiens genomic DNA, chromosome 11q, clone:RP11-121M22,
DEFINITION complete sequence.
```

```
ACCESSION AP002986
VERSION AP002986.2 GI:19263031
KEYWORDS HTG.
```

```
SOURCE Homo sapiens DNA, clone:RP11-121M22.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE 1
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens genomic DNA
Published Only in Database (2000)
```

```
REFERENCE 2 (bases 1 to 157963)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
```

```
JOURNAL Submitted (01-DEC-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(P-mail:hattori@gscl.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Mar 7, 2002 this sequence version replaced gi:11559301.
```

```
FEATURES
source
1..157963
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-121M22"
```

```
BASE COUNT 39742 a 36722 c 37247 g 44252 t
ORIGIN
```

```
Query Match 31.6%; Score 902; DB 9; Length 157963;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 902; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 57 TGAGCCAGAGCGGAGGTAGTCTGCCATCCGACTGGACCCGGACATTAACGCCGCCCG 116
Db 134171 TGAGCCAGAGCGGAGGTAGTCTGCCATCCGACTGGACCCGGACATTAACGCCGCCCG 134230
QY 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGGATCAGGAGTCAATTTTCAGATCACAGC 176
Db 134231 CTACTACTGGCGGGTCCCGAGGACTCCGGGGATCAGGAGTCAATTTTCAGATCACAGC 134290
QY 177 ATTTTCAGAGGACTTTTACCTACACCTGACGCCCGGATGCTCAGTTCTTGGCTCCCGGCTT 236
Db 134291 ATTTTCAGAGGACTTTTACCTACACCTGACGCCCGGATGCTCAGTTCTTGGCTCCCGGCTT 134350
QY 237 CTCCACTCAGCATCTGGCGGTCCCGCTCCAGGGGCTCACCGGGGGCTTTCAGACCTGGG 296
Db 134351 CTCCACTCAGCATCTGGCGGTCCCGCTCCAGGGGCTCACCGGGGGCTTTCAGACCTGGG 134410
QY 297 AGCTGTCTTATTTCTGGGGAGCTGAACGCCCGAGCCGAGTCTCGTCTGCTGTGAGCCT 356
Db 134411 AGCTGTCTTATTTCTGGGGAGCTGAACGCCCGAGCCGAGTCTCGTCTGCTGTGAGCCT 134470
QY 357 GTGGCGGGGCTCCCGGAGCTTTGGCTACCGAGGCGCGAGTATGTCAATAGCCCGCT 416
Db 134471 GTGGCGGGGCTCCCGGAGCTTTGGCTACCGAGGCGCGAGTATGTCAATAGCCCGCT 134530
QY 417 GCCCAATCTAGCGCGCGCGGCGAGCGCAACAGCAGCGGCGCACACCTTCTCCAGCG 476
Db 134531 GCCCAATCTAGCGCGCGCGGCGAGCGCAACAGCAGCGGCGCACACCTTCTCCAGCG 134590
QY 477 CGGGGTGTTCGGCGGGGCTTCCGGAGACCCACCTCTCGCTCGGGGTGGCTCGGG 536
Db 134591 CGGGGTGTTCGGCGGGGCTTCCGGAGACCCACCTCTCGCTCGGGGTGGCTCGGG 134650
QY 537 CTGGAACCCCGCATCTACGGGCGCTTACAGCGCGGCGGGCGGGCTTCGG 596
Db 134651 CTGGAACCCCGCATCTACGGGCGCTTACAGCGCGGCGGGCGGGCTTCGG 134710
QY 597 GGAGAGTCTAGCGCGCGAGTCTGGCGCGCGCAAGCGTTTCTGCTATCCCGCGGTA 656
Db 134711 GGAGAGTCTAGCGCGCGAGTCTGGCGCGCGCAAGCGTTTCTGCTATCCCGCGGTA 134770
QY 657 CGTGAGAGCGTGTGGTGGCGGAGTCAATGTCAAGTTCCACGCGCGGACCTTGA 716
Db 134771 CGTGAGAGCGTGTGGTGGCGGAGTCAATGTCAAGTTCCACGCGCGGACCTTGA 134830
QY 717 ACATTATCTGTGACGCTGTGGCAACGCGCGCGAGTCTTACCGCCATCCAGCATCCT 776
Db 134831 ACATTATCTGTGACGCTGTGGCAACGCGCGCGAGTCTTACCGCCATCCAGCATCCT 134890
QY 777 CAACCCCATCAACATCGTTGGTCAAGGTGCTGCTTCTTAGAGATCGTACTCGGGCC 836
Db 134891 CAACCCCATCAACATCGTTGGTCAAGGTGCTGCTTCTTAGAGATCGTACTCGGGCC 134950
QY 837 CAAGTCAACGGCAATCGCGGCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAAGCT 896
Db 134951 CAAGTCAACGGCAATCGCGGCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAAGCT 135010
QY 897 GAACAAAGTGTAGTACAGACCCCGAGTACTGGGACACTGCCATCTCTTTCACAGGCA 956
Db 135011 GAACAAAGTGTAGTACAGACCCCGAGTACTGGGACACTGCCATCTCTTTCACAGGCA 135070
QY 957 GG 958
Db 135071 GG 135072
```

```
RESULT 5
AC025130/c 172905 bp DNA linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 11 clone RP11-211H6 map 11, WORKING DRAFT
DEFINITION SEQUENCE, 16 unordered pieces.
ACCESSION AC025130
VERSION AC025130.2 GI:7387384
```

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

Human.
Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 172905)

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

TITLE

Homo sapiens chromosome 11, clone RP11-211H6

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 172905)

AUTHORS

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G.,
Campiano, A., Cooke, P., DeArrellano, K., Dewar, K., Collins, S.,
Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Collins, S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

Direct Submission

JOURNAL

Submitted (05-MAR-2000)

REFERENCE

Research, 320 Charles Street, Cambridge, MA 02141, USA

AUTHORS

3 (bases 1 to 172905)

Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G.,
Campiano, A., Cooke, P., DeArrellano, K., Dewar, K., Collins, S.,
Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Collins, S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

Submitted (24-AUG-2002)

JOURNAL

Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Apr 3, 2000 this sequence version replaced qi:7158941.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7808

Center clone name: 211_H_6

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 161899 bases at least Q40

Consensus quality: 167402 bases at least Q30

Consensus quality: 169801 bases at least Q20

Insert size: 177000; agarose-fp

Insert size: 171405; sum-of-contigs

Quality coverage: 4.1 in Q20 bases; agarose-fp

Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 16 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 21: contig of 21 bp in length
 22 121: gap of 100 bp
 122 1343: contig of 1222 bp in length
 1344 1443: gap of 100 bp
 1444 2636: contig of 1193 bp in length
 2637 2736: gap of 100 bp
 2737 4052: contig of 1316 bp in length
 4053 4152: gap of 100 bp
 4153 8044: contig of 3892 bp in length
 8045 8144: gap of 100 bp
 8145 14949: contig of 6805 bp in length
 14950 15049: gap of 100 bp
 15050 23535: contig of 8486 bp in length
 23536 23635: gap of 100 bp
 23636 30261: contig of 6626 bp in length
 30262 30361: gap of 100 bp
 30362 40943: contig of 10580 bp in length
 40942 41041: gap of 100 bp
 41042 54893: contig of 13852 bp in length
 54894 54993: gap of 100 bp
 54994 68888: contig of 13895 bp in length
 68889 68988: gap of 100 bp
 68989 83702: contig of 14714 bp in length
 83703 83802: gap of 100 bp
 83803 98393: contig of 14591 bp in length
 98394 98493: gap of 100 bp
 98494 117356: contig of 18863 bp in length
 117357 117456: gap of 100 bp
 117457 140589: contig of 23133 bp in length
 140590 140689: gap of 100 bp
 140690 172905: contig of 32216 bp in length.

FEATURES

source

1..172905
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11"
 /clone="RP11-211H6"
 /clone_lib="RP11 Human Male BAC"
 1..21
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right
 122..1343
 /note="assembly_fragment"
 1444..2636
 /note="assembly_fragment"
 2737..4052
 /note="assembly_fragment"
 4153..8044
 /note="assembly_fragment"
 8145..14949
 /note="assembly_fragment"
 15050..23535
 /note="assembly_fragment"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

```
misc_feature 23636..30261
/clone="assembly_fragment
/clone_end:SP6
vector_side:right"
30362..40941
/notes="assembly_fragment"
41042..54893
/notes="assembly_fragment"
54994..68888
/notes="assembly_fragment"
68989..83702
/notes="assembly_fragment"
83803..98393
/notes="assembly_fragment"
98494..117356
/notes="assembly_fragment"
117457..140589
/notes="assembly_fragment"
140690..172905
/notes="assembly_fragment"
BASE COUNT 44360 a 41333 c 39516 g 46194 t 1502 others
ORIGIN
Query Match 31.6%; Score 902; DB 2; Length 172905;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 902; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 57 TGACCCAGAGCGGAGGTAGTCGTTCCATCCGACTGGACCCGACATTAACGGCCGCG 116
Db 79527 TGACCCAGAGCGGAGGTAGTCGTTCCATCCGACTGGACCCGACATTAACGGCCGCG 79468;
QY 117 CTACTACTGGGGGTCGCCGAGGACTCCGGGGATCAGGAGACTCATTTTTCAGATCACAGC 176
Db 79467 CTACTACTGGGGGTCGCCGAGGACTCCGGGGATCAGGAGACTCATTTTTCAGATCACAGC 79408
QY 177 ATTTTCAGAGGACTTTTACCTACACCTGACGCGGATGCTCAGTCTTGCTCCCGCCT 236
Db 79407 ATTTTCAGAGGACTTTTACCTACACCTGACGCGGATGCTCAGTCTTGCTCCCGCCT 79348
QY 237 CTCACACTGAGCATCTGGCGCTCCCGCTCCAGGGGCTCACCGGGGCTCTTTACAGACCTCGG 296
Db 79347 CTCACACTGAGCATCTGGCGCTCCCGCTCCAGGGGCTCACCGGGGCTCTTTACAGACCTCGG 79288
QY 297 ACGTGTCTTATTTCTGGGACGTGACGCGGAGCGGAGCTGCTGCTGCTGAGCCT 356
Db 79287 ACGTGTCTTATTTCTGGGACGTGACGCGGAGCGGAGCTGCTGCTGCTGAGCCT 79228
QY 357 GTGCGGGGGCTCCGCGAGCCTTTGGCTACCGAGGCGCGAGTATGCTATTAGCCGCT 416
Db 79227 GTGCGGGGGCTCCGCGAGCCTTTGGCTACCGAGGCGCGAGTATGCTATTAGCCGCT 79168
QY 417 GCCCAATGCTAGCGCGCGCGGCGGCGAGCGCAACAGCCAGGGCGCACACCTTCTCCAGCG 476
Db 79167 GCCCAATGCTAGCGCGCGCGGCGGCGAGCGCAACAGCCAGGGCGCACACCTTCTCCAGCG 79108
QY 477 CCGGGGTGTTCCGGCGGCGCTTCCGAGACCCACCTCTCGTGGGGGTGGCTCGGG 536
Db 79107 CCGGGGTGTTCCGGCGGCGCTTCCGAGAGACCCACCTCTCGTGGGGGTGGCTCGGG 79048
QY 537 CTGGAACCCCGCCATCTACGGGCGCTTGGACCCCTTACAAGCGCGGCGGCGGCTTCGG 596
Db 79047 CTGGAACCCCGCCATCTACGGGCGCTTGGACCCCTTACAAGCGCGGCGGCGGCTTCGG 78988
QY 597 GGAGAGTGTAGCGCGCGCAGGTCTGGGCGGCCAACAGGTTTGTGTCTATCCCGGGTA 656
Db 78987 GGAGAGTGTAGCGCGCGCAGGTCTGGGCGGCCAACAGGTTTGTGTCTATCCCGGGTA 78928
QY 657 CGTGAGACGCTGTGTGCGGAGCAGTCAATGCTCAAGTTCACCGCGCGGACCTGGA 716
Db 78927 CGTGAGACGCTGTGTGCGGAGCAGTCAATGCTCAAGTTCACCGCGCGGACCTGGA 78868
QY 717 ACATTATCTGCTGCTGTGGCAAGCGGCGCGGACTCTACGCCATCCACGATCCT 776
|||||
```

```
Db 78867 ACATTATCTGCTGAGCGCTGTGGCAACGGCGGCGGACTCTACGCCATCCACGATCCT 78808
QY 777 CAACCCCATCAACATCGTTGTGGTCAAGGTGCTGCTTCTTAGAGATCGTGACTCCGGGCC 836
|||||
Db 78807 CAACCCCATCAACATCGTTGTGGTCAAGGTGCTGCTTCTTAGAGATCGTGACTCCGGGCC 78748
QY 837 CAAGGTCAACCGGCAATGCGGCGCTGACGCTGCGCAACTTGTGCTGGCGGAGAGAGCT 896
Db 78747 CAAGGTCAACCGGCAATGCGGCGCTGACGCTGCGCAACTTGTGCTGGCGGAGAGAGCT 78688
QY 897 GAACAAAGTGAAGTACACAGCACCCCGAGTACTGGGACACTGCCATCTCTTCCACGAGCA 956
|||||
Db 78687 GAACAAAGTGAAGTACACAGCACCCCGAGTACTGGGACACTGCCATCTCTTCCACGAGCA 78628
QY 957 GG 958
||
Db 78627 GG 78626

RESULT 6
AC023429
LOCUS
DEFINITION Homo sapiens chromosome 11 clone RP11-121M22, WORKING DRAFT
AC023429
ACCESSION AC023429
VERSION AC023429.19 GI:13569974
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 170682)
Abola,A.P., Bruno,D., Conn,L., Della Rosa,M., Faulkner,D.,
Federici,N., Glukhov,S., Hansen,N., Herman,Z.S., Hymen,R.,
Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
Unpublished
2 (bases 1 to 170682)
Bruno,D., Conn,L., Della Rosa,M., Faulkner,D., Federici,N.,
Glukhov,S., Hansen,N., Hymen,R., Mao,J., Marathe,R.,
Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,
Yu,S. and Davis,R.W.
Direct Submission
Submitted (14-FEB-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
On Apr 10, 2001 this sequence version replaced gi:13562078.
-----Genome Center
Center: Stanford DNA Sequencing and Technology Development
Center
Center code: SDBTDC
Web site: http://sequence-www.stanford.edu/group/human/
Contact: hum-info@sequence.stanford.edu
-----Project Information
Center project name: 837
Center clone name: Rp11-121M22
-----Summary Statistics
Sequencing Vector: M13mp18; X02513
Chemistry: Dye-primer; 12% of reads
Chemistry: Dye-terminator Big Dye; 86% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165770 bases at least Q40
Consensus quality: 167255 bases at least Q30
Consensus quality: 167918 bases at least Q20
Insert size: 172423; agarose-fp
Quality coverage: 9.0x in Q20 bases; agarose-fp
Quality coverage: 9.1x in Q20 bases; sum-of-contigs.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
```



```
|||||
Db 3473 GCATGGCTACAAATTCGTGGTGGCCATCCCGCAGGCGCCTCAAGCATCGACATCCGCCA 3532
QY 2139 GCGCGGTACAAAGGCTGATCGGGATGACAACTACCTGCTCTGAAGACACGCCAAGG 2198
Db 3533 GCGCGGTACAAAGGCTGATCGGGATGACAACTACCTGCTCTGAAGACACGCCAAGG 3592
QY 2199 CAAGTACCTGCTCAACGGGCATTTCTGCTGTGTCGGCGTGGAGGGGACCTGGTGTGA 2258
Db 3593 CAAGTACCTGCTCAACGGGCATTTCTGCTGTGTCGGCGTGGAGGGGACCTGGTGTGA 3652
QY 2259 GGGCAGTCTGCTCGGTACAGCGGCACGGGCACAGCGGTGGAGACCTGTCAGGCTTCCCG 2318
Db 3653 GGGCAGTCTGCTCGGTACAGCGGCACGGGCACAGCGGTGGAGACCTGTCAGGCTTCCCG 3712
QY 2319 GGGCAGTCTGCTCGGTACAGCGGCACGGGCACAGCGGTGGAGACCTGTCAGGCTTCCCG 2378
Db 3713 GGGCAGTCTGCTCGGTACAGCGGCACGGGCACAGCGGTGGAGACCTGTCAGGCTTCCCG 3772
QY 2379 GGTGCGTACTCTTCTATCTGCCCAAGAGCGCTCGGAGGACCAAGTCTCTCATCCCAA 2438
Db 3773 GGTGCGTACTCTTCTATCTGCCCAAGAGCGCTCGGAGGACCAAGTCTCTCATCCCAA 3832
QY 2439 GGACCCCGGGGACCCCTCTGTCTTGCAACAAGCGTCTCTCAGCCCTCTCCAAACAGGTGGA 2498
Db 3833 GGACCCCGGGGACCCCTCTGTCTTGCAACAAGCGTCTCTCAGCCCTCTCCAAACAGGTGGA 3892
QY 2499 GCGCGGAGCAGACGCGCCCTGACGCTGGGTGCTGCGAGCTGGGGCGCTGCTCCGC 2558
Db 3893 GCGCGGAGCAGACGCGCCCTGACGCTGGGTGCTGCGAGCTGGGGCGCTGCTCCGC 3952
QY 2559 GAGCTGCGGAGTGGCTGCAAGAGCGGGGCTGGAGTCCCGGGGCTCCCGCGGCGACGC 2618
Db 3953 GAGCTGCGGAGTGGCTGCAAGAGCGGGGCTGGAGTCCCGGGGCTCCCGCGGCGACGC 4012
QY 2619 CAGGTCCTGCTGTGATGACAGCCATCGGCCCTGTGAGACACAAAGCTCGGGGAGCC 2678
Db 4013 CAGGTCCTGCTGTGATGACAGCCATCGGCCCTGTGAGACACAAAGCTCGGGGAGCC 4072
QY 2679 CTGCCCCACCTGGGAGCTCAGGCGCTGTGCTACCCCTGTCTCAAGAGCTCGGCGGGGATT 2738
Db 4073 CTGCCCCACCTGGGAGCTCAGGCGCTGTGCTACCCCTGTCTCAAGAGCTCGGCGGGGATT 4132
QY 2739 TCAGAGGCGCTCAGTCAAGTGTGGGCGACGAGCGCGGTGCTGCGCCGGGACAGTG 2798
Db 4133 TCAGAGGCGCTCAGTCAAGTGTGGGCGACGAGCGCGGTGCTGCGCCGGGACAGTG 4192
QY 2799 CAAGTGCACCGCAAGCCCGAGGAGCTGACTTCTGCTGAGCGCGTGTGA 2853
Db 4193 CAAGTGCACCGCAAGCCCGAGGAGCTGACTTCTGCTGAGCGCGTGTGA 4247
|||||

RESULT 8
AC025130 172905 bp DNA linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 11 clone RP11-211H6 map 11, WORKING DRAFT
DEFINITION SEQUENCE, 16 unordered pieces.
AC025130
VERSION AC025130.2 GI:7387384
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 172905)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepe,I., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Govette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
2 (bases 1 to 172905)
REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepe,I., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Govette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
JOURNAL
TITLE
AUTHORS
REFERENCE
```

```
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Govette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 172905)
REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G., Collins,S.,
Campopiano,A., Castle,A., Choepe,I., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Govette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 3, 2000 this sequence version replaced gi:7158941.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7808
Center clone name: 211_H_6
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 16189 bases at least Q40
Consensus quality: 167402 bases at least Q30
Consensus quality: 169801 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 171405; sum-of-contigs
Quality coverage: 4.1 in Q20 bases; agarose-fp
Quality coverage: 4.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
```


* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 21: contig of 21 bp in length
22 121: gap of 100 bp
122 1343: contig of 1222 bp in length
1344 1443: gap of 100 bp
1444 2636: contig of 1193 bp in length
2637 2736: gap of 100 bp
2737 4052: contig of 1316 bp in length
4053 4152: gap of 100 bp
4153 8044: contig of 3892 bp in length
8045 8144: gap of 100 bp
8145 14949: contig of 6805 bp in length
14950 15049: gap of 100 bp
15050 23535: contig of 8486 bp in length
23536 23635: gap of 100 bp
23636 30261: contig of 6626 bp in length
30262 30361: gap of 100 bp
30362 40941: contig of 10580 bp in length
40942 41041: gap of 100 bp
41042 54893: contig of 13852 bp in length
54894 54993: gap of 100 bp
54994 68888: contig of 13895 bp in length
68889 68988: gap of 100 bp
68989 83702: contig of 14714 bp in length
83703 83802: gap of 100 bp
83803 98393: contig of 14591 bp in length
98394 98493: gap of 100 bp
98494 117356: contig of 18863 bp in length
117357 117456: gap of 100 bp
117457 140589: contig of 23133 bp in length
140590 140689: gap of 100 bp
140690 172905: contig of 32216 bp in length.
  
```

FEATURES

source

```

1..172905
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-211H6"
/clone_lib="RC1-11 Human Male BAC"

1..21
/feature="assembly_fragment"
clone_end:77
vector_side:right
122..1343
/feature="assembly_fragment"
1444..2636
/feature="assembly_fragment"
2737..4052
/feature="assembly_fragment"
4153..8044
/feature="assembly_fragment"
8145..14949
/feature="assembly_fragment"
15050..23535
/feature="assembly_fragment"
23636..30261
/feature="assembly_fragment"
clone_end:SP6
vector_side:right
30362..40941
/feature="assembly_fragment"
41042..54893
/feature="assembly_fragment"
54994..68888
/feature="assembly_fragment"
68989..83702
/feature="assembly_fragment"
83803..98393
/feature="assembly_fragment"
  
```

```

misc_feature      98494..117356
                  /note="assembly_fragment"
misc_feature      117457..140589
                  /note="assembly_fragment"
misc_feature      140590..172905
                  /note="assembly_fragment"

BASE COUNT      44360 a 41333 c 39516 g 46194 t 1502 others
ORIGIN

Query Match      12.5% Score 358; DB 2; Length 172905;
Best Local Similarity 100.0%; Pred. No. 2.1e-183;
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2079 GCATGGCTACAAATTCGTGGTGGCCATCCCGCAGGCGCCTCAAGCATCCACATCCGCCA 2138
          |||||||
Db 54519 GCATGGCTACAAATTCGTGGTGGCCATCCCGCAGGCGCCTCAAGCATCCACATCCGCCA 54578

Qy 2139 GCGCGGTTACAAAGGCTGATCGGGGATGACAACTACTCTGGCTCTGAAGACACA:CCCAAGG 2198
          |||||||
Db 54579 GCGCGGTTACAAAGGCTGATCGGGGATGACAACTACTCTGGCTCTGAAGACACA:CCCAAGG 54638

Qy 2199 CAAGTACCTGCTCAACGGGCATTTCGTGGTCTCGGCGGTGGAGCGGACCTGGTGGTAA 2258
          |||||||
Db 54639 CAAGTACCTGCTCAACGGGCATTTCGTGGTCTCGGCGGTGGAGCGGACCTGGTGGTAA 54698

Qy 2259 GGGCAGTCTGCTGGGTACAGCGGCACAGCGGTGGAGAGCTTCAGGCTTCGCC 2318
          |||||||
Db 54699 GGGCAGTCTGCTGGGTACAGCGGCACAGCGGTGGAGAGCTTCAGGCTTCGCC 54758

Qy 2319 GCCCATCTCGGAGCGGTGACCGTGGAGGTCTCTCGGTGGGGAAGATGACACGCCGCCG 2378
          |||||||
Db 54759 GCCCATCTCGGAGCGGTGACCGTGGAGGTCTCTCGGTGGGGAAGATGACACGCCGCCG 54818

Qy 2379 GGTCCGTACTCTCTCTATCTGCCCCAAGAGCCTCGGAGGACAAAGTCTCTCATCCC 2436
          |||||||
Db 54819 GGTCCGTACTCTCTCTATCTGCCCCAAGAGCCTCGGAGGACAAAGTCTCTCATCCC 54876
  
```

RESULT 9
 AC023429/c

LOCUS Homo sapiens chromosome 11 clone RP11-121M22, WORKING DRAFT
 DEFINITION SEQUENCE, 7 unordered pieces.

AC023429
 AC023429.19 GI:13569974
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS Homo sapiens.
 SOURCE Homo sapiens

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 170682)
 Abola, A.P., Bruno, D., Conn, L., Dela Rosa, M., Faulkner, D.,
 Federspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,
 Mao, J., Komp, C., Kottler, S., Lam, B., Marathe, R., Miranda, M.,
 Morehouse, A.J., Nguyen, M., Oefner, P., Palm, C.J., Ramirez, D.,
 Southwick, A.M., Webb, C., Wilhelmy, J., Yu, S. and Davis, R.W.
 Unpublished

JOURNAL

2 (bases 1 to 170682)
 Bruno, D., Conn, L., Dela Rosa, M., Faulkner, D., Federspiel, N.,
 Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R.,
 Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J.,
 Yu, S. and Davis, R.W.
 Direct Submission

AUTHORS

Submitted (14-FEB-2000) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 On Apr 10, 2001 this sequence version replaced gi:13562078.

COMMENT

Center: Stanford DNA Sequencing and Technology Development
 Center
 Center code: SDSTDC
 Web site: <http://sequence-www.stanford.edu/group/human/>
 Contact: hum-info@sequence.stanford.edu

```
----- Project Information
Center project name: 837
Center clone name: Rpl1-121M22
----- Summary Statistics
Sequencing Vector: M13mp18; X02513
Chemistry: Dye-primer; 12% of reads
Chemistry: Dye-terminator Big Dye; 86% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165770 bases at least Q40
Consensus quality: 167255 bases at least Q30
Consensus quality: 167918 bases at least Q20
Insert size: 172423; agarose-fp
Insert size: 170082; sum-of-contigs
Quality coverage: 9.0x in Q20 bases; agarose-fp
Quality coverage: 9.1x in Q20 bases; sum-of-contigs
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1856: contig of 1856 bp in length
* 1857 1956: gap of unknown length
* 1957 7065: contig of 5109 bp in length
* 7066 7165: gap of unknown length
* 7166 21089: contig of 13924 bp in length
* 21090 21189: gap of unknown length
* 21190 34145: contig of 12956 bp in length
* 34146 34246: gap of unknown length
* 34246 48562: contig of 14317 bp in length
* 48563 48662: gap of unknown length
* 48663 109179: contig of 60517 bp in length
* 109180 109279: gap of unknown length
* 109280 170682: contig of 61403 bp in length.
FEATURES
    source
        1..170682
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="11"
            /clone="Rpl1-121M22"
            /clone_lib="RPC1 human BAC library 11"
        1..1856
            /note="assembly_name:Contig25"
            1957..7065
            /note="assembly_name:Contig26"
            7166..21089
            /note="assembly_name:Contig27"
            21190..34145
            /note="assembly_name:Contig28"
            34246..48562
            /note="assembly_name:Contig29"
            clone_end:SP6"
            48663..109179
            /note="assembly_name:Contig30"
            109280..170682
            /note="assembly_name:Contig31"
            clone_end:7"
BASE COUNT 43947 a 39732 c 39514 g 46884 t 605 others
ORIGIN
Query Match 10.1%; Score 288; DB 2; Length 170682;
Best Local Similarity 100.0%; Pred. No. 3.4e-145;
Matches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1258 GGTGACTGCTCTGACCAACCCAGCAGCCATCTCCCTGCCGAGATCTGCCGGGC 1317
|
|
|
Db 44571 GGTGACTGCTCTGACCAACCCAGCAGCCATCTCCCTGCCGAGATCTGCCGGGC 44512
|
|
|
QY 1318 GCCAGCTACACCTGAGCCAGCAGTCCGAGCTGGCTTTTGGCTGGGCTCCAGCCCTGT 1377
|
|
|
Db 44511 GCCAGCTACACCTGAGCCAGCAGTCCGAGCTGGCTTTTGGCTGGGCTCCAGCCCTGT 44552
|
|
|
```

```
QY 1378 CCTTACATGCAGTACTGCACCAAGCTGTGTGTGCACCGGAAGGCCAAGGACAGATGGTG 1437
|
|
|
Db 44451 CCTTACATGCAGTACTGCACCAAGCTGTGTGTGCACCGGAAGGCCAAGGACAGATGGTG 44392
|
|
|
QY 1438 TGCAGAGCCGCCACTCTCCCTGGGCGGATGGCACCAGCTGTGGCGAGGCGAGCTCTGC 1497
|
|
|
Db 44391 TGCAGAGCCGCCACTCTCCCTGGGCGGATGGCACCAGCTGTGGCGAGGCGAGCTCTGC 44332
|
|
|
QY 1498 CTCAAGGGCCCTGCGTGGAGACACAACCTCAACAAGCACACAGGGGTG 1545
|
|
|
Db 44331 CTCAAGGGCCCTGCGTGGAGACACAACCTCAACAAGCACACAGGGGTG 44284
|
|
|
RESULT 10
AC101990 182656 bp DNA linear HTG 21-AUG-2002
LOCUS Mus musculus clone RP24-371J2, WORKING DRAFT SEQUENCE, 16 unordered
DEFINITION pieces.
AC101990
VERSION AC101990.2 GI:22381363
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 182656)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP24-371J2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 182656)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferrelira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliiev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., MCPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phukhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 182656)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepe,Y., Collymore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferrelira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phukhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
```

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17060766.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L17833
Center clone name: 37L_J2

Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 178030 bases at least Q40
Consensus quality: 179964 bases at least Q30
Consensus quality: 180647 bases at least Q20
Insert size: 172000; agarose-fp
Insert size: 181156; sum-of-contigs
Quality coverage: 8.1 in Q20 bases; agarose-fp
Quality coverage: 7.7 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 361: contig of 361 bp in length
362 461: gap of 100 bp
462 1538: contig of 1077 bp in length
1539 1638: gap of 100 bp
1639 3397: contig of 1759 bp in length
3398 3497: gap of 100 bp
3498 4947: contig of 1450 bp in length
4948 5047: gap of 100 bp
5048 6893: contig of 1846 bp in length
6894 6993: gap of 100 bp
6994 10395: contig of 3402 bp in length
10396 10495: gap of 100 bp
10496 12584: contig of 2089 bp in length
12585 12684: gap of 100 bp
12685 14977: contig of 2293 bp in length
14978 15077: gap of 100 bp
15078 18181: contig of 3104 bp in length
18182 18281: gap of 100 bp
18282 22654: contig of 4373 bp in length
22655 22754: gap of 100 bp
22755 31261: contig of 8507 bp in length
31262 31361: gap of 100 bp
31362 43587: contig of 12226 bp in length
43588 43687: gap of 100 bp
43688 57612: contig of 13925 bp in length
57613 57712: gap of 100 bp
57713 77467: contig of 19755 bp in length
77468 77567: gap of 100 bp
77568 108724: contig of 31157 bp in length
108725 108824: gap of 100 bp
108825 182656: contig of 73832 bp in length.

Location/Qualifiers
1. 182656
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-37LJ2"

misc_feature
1. 361
/clone_lib="RPC1-24 Male Mouse BAC"
/note="assembly_fragment"
462. 1538
/note="assembly_fragment"
1639. 3397
/note="assembly_fragment"
3498. 4947
/note="assembly_fragment"
5048. 6893
/note="assembly_fragment"
6994. 10395
/note="assembly_fragment"
10496. 12584
/note="assembly_fragment"
12685. 14977
/note="assembly_fragment"
15078. 18181
/note="assembly_fragment"
18282. 22654
/note="assembly_fragment"
22755. 31261
/note="assembly_fragment"
31362. 43587
/note="assembly_fragment"
43688. 57612
/note="assembly_fragment"
57713. 77467
/note="assembly_fragment"
77568. 108724
/note="assembly_fragment"
108825. 182656
/note="assembly_fragment"
BASE COUNT 51468 a 41160 c 39770 g 48758 t 1500 other:s
ORIGIN

Query Match 2.6%; Score 74; DB 2; Length 182656;
Best Local Similarity 100.0%; Pred. No. 2.4e-28;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1009 ACCATGTGTGACCCCAAGAGAGCTGCTGTCTATTGAGGACGATGGCTTCCATCAGCC 1068
|||||
Db 16446 ACCATGTGTGACCCCAAGAGAGCTGCTGTCTATTGAGGACGATGGCTTCCATCAGCC 16505
|||||
QY 1069 TTCACCACTGCCCA 1082
|||||
Db 16506 TTCACCACTGCCCA 16519

RESULT 11
AC126507/c 178764 bp DNA linear HTG 24-JUL-2002
LOCUS Rattus norvegicus clone CH230-254N12, *** SEQUENCING IN PROGRESS
DEFINITION *** 49 unordered pieces.

ACCESSION AC126507
VERSION AC126507.1 GI:21700463
KEYWORDS HTG; HTGS PHASE1.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 178764)
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, T., Barbaria, J., Benton, J., Blin, K., Blankenburg, K., Bonin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C., Burck, P., Burckett, C., Burrell, K.L., Byrd, K.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,

FEATURES
source

Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Correll, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Jacobson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Lucier, R., Luna, R., Ma, J.,
 Lozadó, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
 Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczkyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 178764)
 Worley, K.C.
 Direct Submission
 Submitted (06-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 178764)
 Worley, K.C.
 Direct Submission
 Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GZFU
 Center clone name: CH230-254N12
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.950329
 Consensus quality: 124573 bases at least Q40
 Consensus quality: 131850 bases at least Q30
 Consensus quality: 137700 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 49 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1037: contig of 1037 bp in length
 * 1038 1137: gap of unknown length
 * 1138 2250: contig of 1113 bp in length
 * 2251 3579: gap of unknown length
 * 3579: contig of 1229 bp in length
 * 3679: gap of unknown length
 * 4874: contig of 1195 bp in length
 * 4875: gap of unknown length
 * 4975: contig of 1551 bp in length
 * 6525: gap of unknown length
 * 6625: contig of 1280 bp in length
 * 7905: gap of unknown length
 * 8005: contig of 1719 bp in length
 * 9724: gap of unknown length
 * 9725: contig of 1618 bp in length
 * 9825: gap of unknown length
 * 11442: contig of 1358 bp in length
 * 11543: gap of unknown length
 * 12900: contig of 1150 bp in length
 * 13001: gap of unknown length
 * 14150: contig of 1315 bp in length
 * 14250: gap of unknown length
 * 15565: contig of 1810 bp in length
 * 15665: gap of unknown length
 * 17475: contig of 1115 bp in length
 * 17575: gap of unknown length
 * 18690: contig of 1479 bp in length
 * 18790: gap of unknown length
 * 20269: contig of 1558 bp in length
 * 20369: gap of unknown length
 * 21927: contig of 1971 bp in length
 * 22027: gap of unknown length
 * 23998: contig of 1581 bp in length
 * 24098: gap of unknown length
 * 25679: contig of 1649 bp in length
 * 25779: gap of unknown length
 * 27428: contig of 2149 bp in length
 * 27528: gap of unknown length
 * 29677: contig of 1100 bp in length
 * 29777: gap of unknown length
 * 30877: gap of unknown length
 * 30977: contig of 1854 bp in length
 * 32831: gap of unknown length
 * 32832: contig of 2488 bp in length
 * 32932: gap of unknown length
 * 35420: contig of 1874 bp in length
 * 35519: gap of unknown length
 * 37393: contig of 2342 bp in length
 * 37493: gap of unknown length
 * 37934: contig of 2854 bp in length
 * 37949: gap of unknown length
 * 39835: contig of 3757 bp in length
 * 39836: gap of unknown length
 * 42789: contig of 2442 bp in length
 * 42890: gap of unknown length
 * 46647: contig of 4517 bp in length
 * 46747: gap of unknown length
 * 49189: contig of 4088 bp in length
 * 49289: gap of unknown length
 * 53805: contig of 3038 bp in length
 * 53905: gap of unknown length
 * 56202: contig of 4108 bp in length
 * 56302: gap of unknown length
 * 58719: contig of 2979 bp in length
 * 58819: gap of unknown length
 * 58820: contig of 4085 bp in length
 * 62904: gap of unknown length
 * 63004: contig of 2928 bp in length
 * 65932: gap of unknown length
 * 65933: contig of 4088 bp in length
 * 66033: gap of unknown length
 * 70120: contig of 3038 bp in length
 * 70121: gap of unknown length
 * 70220: contig of 4108 bp in length
 * 73258: gap of unknown length
 * 73259: contig of 2979 bp in length
 * 73359: gap of unknown length
 * 77466: contig of 5342 bp in length
 * 77467: gap of unknown length
 * 80545: contig of 4821 bp in length
 * 80546: gap of unknown length
 * 80646: contig of 5342 bp in length
 * 85987: gap of unknown length
 * 85988: contig of 4821 bp in length
 * 86088: gap of unknown length
 * 90908: contig of 4821 bp in length
 * 91008: gap of unknown length
 * 90909

```

* 91009 97918: contig of 6910 bp in length
* 97919 98018: gap of unknown length
* 98019 103502: contig of 5484 bp in length
* 103503 112670: gap of unknown length
* 103603 112670: contig of 8968 bp in length
* 112571 112670: gap of unknown length
* 112671 117924: contig of 5254 bp in length
* 117925 118024: gap of unknown length
* 118025 125019: contig of 6995 bp in length
* 125020 125119: gap of unknown length
* 125120 134087: contig of 8968 bp in length
* 134088 134187: gap of unknown length
* 134188 143851: contig of 9664 bp in length
* 143852 143951: gap of unknown length
* 143952 151699: contig of 7748 bp in length
* 151700 151799: gap of unknown length
* 151800 158828: contig of 7027 bp in length
* 158827 158928: gap of unknown length
* 158927 168522: contig of 9596 bp in length
* 168523 168623: gap of unknown length
* 168623 178764: contig of 10142 bp in length.

FEATURES
    source
    Location/Qualifiers
    1..178764
    /organism="Rattus norvegicus"
    /db_xref="taxon:10116"
    /clone="CH230-254N12"

BASE COUNT 47487 a 39119 c 38897 g 45423 t 7838 others
ORIGIN

Query Match      2.1%  Score 61;  DB 2;  Length 178764;
Best Local Similarity 100.0%;  Pred. No. 3e-21;  0;  Gaps 0;
Matches 61;  Conservative 0;  Mismatches 0;  Indels 0;

Qy 1942 GTCTGTGTCGAAGTCATCAAGCTGGCTGTGATGGGAACCTGGCTCCAGAAG 2001
|||||
Db 150003 GTCTGTGTCGAAGTCATCAAGCTGGCTGTGATGGGAACCTGGCTCCAGAAG 149944

Qy 2002 A 2002
Db 149943 A 149943

RESULT 12
AC126507
LOCUS
DEFINITION Rattus norvegicus clone CH230-254N12, *** SEQUENCING IN PROGRESS.
AC126507
ACCESSION AC126507.1 GI:21700463
VERSION
KEYWORDS HTG; HTGS_PHASE1.
SOURCE
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 178764)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbacia,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,K., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franta,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,

```

```

Homsí, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, F., Johnson, R., Jollivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, C., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morjan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Oraguñe, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
Peters, L., Pickens, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,
Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 178764)
Worley, K.C.
Direct Submission
Submitted (06-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 178764)
Worley, K.C.
Direct Submission
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
-----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Center project name: GZFU
Center clone name: CH230-254N12
-----
Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 124573 bases at least Q40
Consensus quality: 131850 bases at least Q30
Consensus quality: 137700 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1037: contig of 1037 bp in length
* 1137: gap of unknown length
* 1138 1138: contig of 1113 bp in length
* 2250: gap of unknown length
* 2350: gap of unknown length
* 2351 2351: contig of 1229 bp in length
* 3579: gap of unknown length
* 3679: gap of unknown length
* 3680 3680: contig of 1195 bp in length
* 4875 4875: gap of unknown length
* 4974: gap of unknown length
* 5625: contig of 1551 bp in length
* 6625: gap of unknown length

```

* 6626	7905:	contig of 1280 bp in length	* 117925	118024:	gap of unknown length
* 7906	8005:	gap of unknown length	* 118025	125019:	contig of 6995 bp in length
* 8006	9724:	contig of 1719 bp in length	* 125020	125119:	gap of unknown length
* 9725	9824:	gap of unknown length	* 125120	134087:	contig of 8968 bp in length
* 9825	11442:	contig of 1618 bp in length	* 134088	134187:	gap of unknown length
* 11443	11542:	gap of unknown length	* 134188	143851:	contig of 9664 bp in length
* 11543	12900:	contig of 1358 bp in length	* 143852	143951:	gap of unknown length
* 12901	13000:	gap of unknown length	* 143952	151699:	contig of 7748 bp in length
* 13001	14150:	contig of 1150 bp in length	* 151700	151799:	gap of unknown length
* 14151	14250:	gap of unknown length	* 151800	158826:	contig of 7027 bp in length
* 14251	15365:	contig of 1315 bp in length	* 158827	158925:	gap of unknown length
* 15366	15665:	gap of unknown length	* 158927	168522:	contig of 9596 bp in length
* 15666	17475:	contig of 1810 bp in length	* 168523	168622:	gap of unknown length
* 17476	17575:	gap of unknown length	* 168623	178764:	contig of 10142 bp in length.
* 17576	18690:	contig of 1115 bp in length	FEATURES	Location/Qualifiers	
* 18691	18790:	gap of unknown length	1. 178764		
* 18791	20269:	contig of 1479 bp in length	/organism="Rattus norvegicus"		
* 20270	20369:	gap of unknown length	/db_xref="taxon:10116"		
* 20370	21927:	contig of 1558 bp in length	/clone="CH230-254N12"		
* 21928	22027:	gap of unknown length	BASE COUNT 47487 a 39119 c 38897 g 45423 t 7838 others		
* 22028	23998:	contig of 1971 bp in length	ORIGIN		
* 23999	24098:	gap of unknown length	Query Match 1.5% Score 44: DB 2: Length 178764;		
* 24099	25679:	contig of 1581 bp in length	Best Local Similarity 100.0%; Pred. No. 5.7e-12; Length 178764;		
* 25680	25779:	gap of unknown length	Matches 44: Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
* 25780	27428:	contig of 1649 bp in length			
* 27429	27528:	gap of unknown length			
* 27529	29677:	contig of 2149 bp in length			
* 29678	29777:	gap of unknown length			
* 29778	30977:	contig of 1100 bp in length			
* 30978	30977:	gap of unknown length			
* 30978	32831:	contig of 1854 bp in length			
* 32832	32931:	gap of unknown length			
* 32932	35419:	contig of 2488 bp in length			
* 35420	35519:	gap of unknown length			
* 35520	37393:	contig of 1874 bp in length			
* 37394	37433:	gap of unknown length			
* 37494	39835:	contig of 2342 bp in length			
* 39836	39935:	gap of unknown length			
* 39936	42789:	contig of 2854 bp in length			
* 42790	42889:	gap of unknown length			
* 42890	46646:	contig of 3757 bp in length			
* 46647	46746:	gap of unknown length			
* 46747	49188:	contig of 2442 bp in length			
* 49189	49288:	gap of unknown length			
* 49289	53805:	contig of 4517 bp in length			
* 53806	53905:	gap of unknown length			
* 53906	56202:	contig of 2297 bp in length			
* 56203	56302:	gap of unknown length			
* 56303	58719:	contig of 2417 bp in length			
* 58720	58819:	gap of unknown length			
* 58820	62904:	contig of 4085 bp in length			
* 62905	63004:	gap of unknown length			
* 63005	65932:	contig of 2928 bp in length			
* 65933	66032:	gap of unknown length			
* 66033	70130:	contig of 4088 bp in length			
* 70130	70220:	gap of unknown length			
* 70221	73258:	contig of 3038 bp in length			
* 73259	73358:	gap of unknown length			
* 73359	77466:	contig of 4108 bp in length			
* 77467	77566:	gap of unknown length			
* 77567	80545:	contig of 2979 bp in length			
* 80546	80645:	gap of unknown length			
* 80646	85987:	contig of 5342 bp in length			
* 85988	86087:	gap of unknown length			
* 86088	90908:	contig of 4821 bp in length			
* 90909	91008:	gap of unknown length			
* 91009	97918:	contig of 6910 bp in length			
* 97919	98018:	gap of unknown length			
* 98019	103502:	contig of 5484 bp in length			
* 103503	103603:	gap of unknown length			
* 103603	112570:	contig of 8968 bp in length			
* 112571	112670:	gap of unknown length			
* 112671	117924:	contig of 5254 bp in length			

AC101990	182656 bp	DNA	linear	HTG 21-AUG-2002
LOCUS				
DEFINITION	Mus musculus clone RP24-371J2, WORKING DRAFT SEQUENCE, 16 unordered pieces.			
ACCESSION	AC101990			
VERSION	AC101990.2 GI:22381363			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	1 (bases 1 to 182656)			
AUTHORS	Birren,B., Nusbaum,C. and Lander,E.			
TITLE	Mus musculus, clone RP24-371J2			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 182656)			
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,W., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C., Kanat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,I., Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Melidrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.			
Direct Submission				
Submitted (23-NOV-2001)	Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA			
JOURNAL				

REFERENCE
AUTHORS

3 (bases 1 to 182656)
 Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
 Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
 Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
 McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
 Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL

Direct Submission
 Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Aug 21, 2002 this sequence version replaced gi:17060766.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L17833
 Center Clone name: 371_J_2
 ----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 178030 bases at least Q40
 Consensus quality: 179964 bases at least Q30
 Consensus quality: 180647 bases at least Q20
 Insert size: 172000; agarose-ff
 Insert size: 181156; sum-of-contigs
 Quality coverage: 8.1 in Q20 bases; agarose-ff
 Quality coverage: 7.7 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 16 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 361: contig of 361 bp in length
 * 362 461: gap of 100 bp
 * 462 1538: contig of 1077 bp in length
 * 1539 1638: gap of 100 bp
 * 1639 3397: contig of 1759 bp in length
 * 3398 3497: gap of 100 bp
 * 3498 4947: contig of 1450 bp in length
 * 4948 5047: gap of 100 bp
 * 5048 6893: contig of 1845 bp in length
 * 6894 6993: gap of 100 bp
 * 6994 10395: contig of 3402 bp in length
 * 10396 10495: gap of 100 bp
 * 10496 12584: contig of 2089 bp in length
 * 12585 12684: gap of 100 bp
 * 12685 14977: contig of 2293 bp in length
 * 14978 15077: gap of 100 bp
 * 15078 18181: contig of 3104 bp in length
 * 18182 18281: gap of 100 bp
 * 18282 22654: contig of 4373 bp in length

* 22655 22754: gap of 100 bp
 * 22755 31261: contig of 8507 bp in length
 * 31262 31361: gap of 100 bp
 * 31362 43587: contig of 12226 bp in length
 * 43588 43687: gap of 100 bp
 * 43688 57612: contig of 13925 bp in length
 * 57613 57712: gap of 100 bp
 * 57713 77467: contig of 19755 bp in length
 * 77468 77567: gap of 100 bp
 * 77568 108724: contig of 31157 bp in length
 * 108725 108824: gap of 100 bp
 * 108825 182656: contig of 73832 bp in length.

FEATURES

Location/Qualifiers
 1..182656
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP24-371J2"
 /clone_lib="RPC1-24 Male Mouse BAC"
 1..361
 /note="assembly_fragment"

misc_feature

462..1538
 /note="assembly_fragment"

misc_feature

1639..3397
 /note="assembly_fragment"

misc_feature

3498..4947
 /note="assembly_fragment"

misc_feature

5048..6893
 /note="assembly_fragment"

misc_feature

6994..10395
 /note="assembly_fragment"

misc_feature

10496..12584
 /note="assembly_fragment"

misc_feature

12685..14977
 /note="assembly_fragment"

misc_feature

15078..18181
 /note="assembly_fragment"

misc_feature

18282..22654
 /note="assembly_fragment"

misc_feature

22755..31261
 /note="assembly_fragment"

misc_feature

31362..43587
 /note="assembly_fragment"

misc_feature

43688..57612
 /note="assembly_fragment"

misc_feature

57713..77467
 /note="assembly_fragment"

misc_feature

77568..108724
 /note="assembly_fragment"

misc_feature

108825..182656
 /note="assembly_fragment"

BASE COUNT 51468 a 41160 c 39770 g 48758 t 1500 others

ORIGIN

Query Match 1.3%; Score 36; DB 2; Length 182656;
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 449 ACAGCCAGGGCGCACACCTTCTCCAGCGCGGGTG 484

DB 39334 ACAGCCAGGGCGCACACCTTCTCCAGCGCGGGTG 39299

RESULT 14

BC009667

LOCUS

DEFINITION

Mus musculus, similar to a disintegrin and metalloproteinase with
 thrombospondin motifs 1 (ADAMTS-1), clone IMAGE:3491991, mRNA,

partial cds.

ACCESSION

BC009667

VERSION

BC009667.1

KEYWORDS

SOURCE

ORGANISM

BC009667 3927 bp mRNA linear ROD 07-AUG-2002

Mus musculus, similar to a disintegrin and metalloproteinase with

thrombospondin motifs 1 (ADAMTS-1), clone IMAGE:3491991, mRNA,

partial cds.

ACCESSION BC009667

VERSION BC009667.1

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Strausberg, R.

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (29-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk
Email: cgabs-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRK Plate: 8 Row: J Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES
source

Location/Qualifiers
1. 3927
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="CZECH II"
/clone="IMAGE:3491991"
/tissue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMV."
/clone_lib="NCL_CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
cl. 1024
/codon_start=2
/product="similar to a disintegrin and metalloproteinase with thrombospondin motifs 1 (ADAMTS-1)"
/protein_id="AAH09667.1"
/db_xref="GI:16307156"
translation="PRVRKCKLICRANGTYFYVLPAPRVVDGTLCTPDSTVCVQGRIC
IKAGCDGNLGGSKKFKDKCGVCGDNKCKRVTLGFTKPMHGYFNFWAIPAGASSIDIR
ORGYKGLIGDDNLYLAKNSQGYLLNGHFVSAVERDLVKGSLVLRSGTCTAVESLQ
ASRPLEPLTEVLSVCKMTPTPRVYSFYLKPEPRDKSTRPKDPRGSLVLRNLSVL
SNQVQPNRPPARVAGSWGFCVSCGSLQKRAVDCRDSPGQGGACACDVDRHPL
KRACEPCTWELGNWSPKSCGGRGFRKPLKCVGHGRLLRDQCDLRKRPQELDF
CVLRPC"

BASE COUNT 938 a 920 c 1093 g 976 t

ORIGIN

Query Match 1.2%; Score 35; DB 10; Length 3927;
Best Local Similarity 100.0%; Pred. No. 6.4e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1942 GTCTGTGTCACGAGGCAATGTCATCAAGCTGGCTG 1976
|||||
Db 113 GTCTGTGTCACGAGGCAATGTCATCAAGCTGGCTG 147

RESULT 15
AC099229
LOCUS
DEFINITION
AC099229
AC099229.3 GI:21736081
KEYWORDS
HTG; HTGS_PHASE1.

AC099229
Rattus norvegicus clone CH230-3019, *** SEQUENCING IN PROGRESS ***,
48 unordered pieces.
AC099229
AC099229.3 GI:21736081
HTG; HTGS_PHASE1.

SOURCE
ORGANISM

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 86229)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alzubrooks, S.L., Amaral, H.C., Are, J.R., Avele, M., Banks, T., Barabara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chaves, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Sisson, I., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sison, I., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, S., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 86229)

Worley, K.C.

REFERENCE
AUTHORS

Direct Submission

Submitted (09-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 86229)

Worley, K.C.

REFERENCE
AUTHORS

Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 12, 2002 this sequence version replaced gi:17973111.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GHDX

Center clone name: CH230-3019

----- Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 60469 bases at least Q40
Consensus quality: 63966 bases at least Q30
Consensus quality: 66267 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1289 1288: contig of 1288 bp in length
1389 1388: gap of unknown length
1389 2502: contig of 1114 bp in length
2503 2602: gap of unknown length
2603 3645: contig of 1043 bp in length
3646 3745: gap of unknown length
3746 4866: contig of 1121 bp in length
4867 4966: gap of unknown length
4967 6411: contig of 1445 bp in length
6412 6511: gap of unknown length
6512 7670: contig of 1159 bp in length
7671 7770: gap of unknown length
7771 8823: contig of 1053 bp in length
8824 8923: gap of unknown length
8924 10105: contig of 1182 bp in length
10106 10205: gap of unknown length
10206 11283: contig of 1078 bp in length
11284 11383: gap of unknown length
11384 12892: contig of 1509 bp in length
12893 12992: gap of unknown length
12993 15116: contig of 2124 bp in length
15117 15216: gap of unknown length
15217 16395: contig of 1179 bp in length
16396 16495: gap of unknown length
16496 17597: contig of 1102 bp in length
17598 17697: gap of unknown length
17698 18822: contig of 1125 bp in length
18823 18922: gap of unknown length
18923 20128: contig of 1206 bp in length
20129 20228: gap of unknown length
20229 22372: contig of 2144 bp in length
22373 22472: gap of unknown length
22473 23954: contig of 1482 bp in length
23955 24054: gap of unknown length
24055 25745: contig of 1691 bp in length
25746 25845: gap of unknown length
25846 27386: contig of 1541 bp in length
27387 27486: gap of unknown length
27487 28937: contig of 1451 bp in length
28938 29037: gap of unknown length
29039 30661: contig of 1024 bp in length
30662 30161: gap of unknown length
30162 31758: contig of 1597 bp in length
31759 31858: gap of unknown length
31859 33015: contig of 1157 bp in length
33016 33115: gap of unknown length
33116 34336: contig of 1421 bp in length
34337 34636: gap of unknown length
34637 35951: contig of 1315 bp in length
35952 36051: gap of unknown length
36052 37516: contig of 1465 bp in length
37517 37616: gap of unknown length
37617 39220: contig of 1604 bp in length
39221 39320: gap of unknown length
39321 40785: contig of 1465 bp in length
40786 40885: gap of unknown length
40886 42770: contig of 1885 bp in length
42771 42870: gap of unknown length
42872 45007: contig of 2137 bp in length
45008

45107: gap of unknown length
45108 46717: contig of 1610 bp in length
46718 46817: gap of unknown length
46819 48717: contig of 1900 bp in length
48718 48817: gap of unknown length
48819 50105: contig of 1288 bp in length
50106 50205: gap of unknown length
50206 51842: contig of 1637 bp in length
51843 51943: gap of unknown length
51944 53289: contig of 1347 bp in length
53290 53389: gap of unknown length
53390 54765: contig of 1376 bp in length
54766 54865: gap of unknown length
54866 58044: contig of 3179 bp in length
58045 58144: gap of unknown length
58145 60076: contig of 1932 bp in length
60077 60176: gap of unknown length
60177 62070: contig of 1894 bp in length
62071 62170: gap of unknown length
62171 64612: contig of 2442 bp in length
64613 64713: gap of unknown length
64714 67144: contig of 2432 bp in length
67145 67244: gap of unknown length
67245 69807: contig of 2563 bp in length
69808 69907: gap of unknown length
69909 71762: contig of 1855 bp in length
71763 71862: gap of unknown length
71863 74111: contig of 2249 bp in length
74112 74211: gap of unknown length
74212 76217: contig of 2006 bp in length
76218 76317: gap of unknown length
76319 78125: contig of 1808 bp in length
78126 78225: gap of unknown length
78226 81630: contig of 3405 bp in length
81631 81730: gap of unknown length
81731 86229: contig of 4499 bp in length.

FEATURES

source

1. 86229
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-3019"

BASE COUNT 23125 a 17965 c 17868 g 22544 t 4727 others
ORIGIN

Query Match 0.8%; Score 23; DB 2; Length 86229;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2008 GACAAGTGTGGGTGTGTGGGG 2030
|||||
Db 64987 GACAAGTGTGGGTGTGTGGGG 65009

Search completed: May 2, 2003, 04:44:18
Job time : 12343 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2003, 22:50:37 ; Search time 606 Seconds
(without alignments)
10602.230 Million cell updates/sec

Title: US-10-009-332-2
Perfect score: 2853
Sequence: 1 atgcttttgcgggcatcctt.....gcgtcctgagccgtgctga-2853

Scoring table: OLIGO.NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :	N_Geneseq_101002:*
1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24:	/SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2853	100.0	2853	22	AAH41003 Human metalloprote
2	2695	94.5	2853	24	AAD35569 Human protease cDN
3	2695	94.5	3446	24	AAD35571 Human protease cDN
4	2089	73.2	2930	24	ABK12894 Human protease PRT
5	2089	73.2	2937	24	AAS97182 Human metalloprote
6	1035	36.3	1104	24	AAD35570 Human protease cDN
7	903	31.7	966	24	AAD35568 Human protease cDN
8	481	16.9	1143	21	AAA95831 Human metalloprote
9	50	1.8	1518	21	AAA95827 Rat metalloprotein

c	10	30	1.1	30	22	AAH41029	Adaptor primer SEQ
c	11	30	1.1	30	22	AAH41030	Adaptor primer SEQ
c	12	30	1.1	41	22	AAH41017	PCR primer specifi
c	13	30	1.1	42	22	AAH41016	PCR primer specifi
c	14	27	0.9	27	22	AAH41018	PCR primer specifi
c	15	27	0.9	27	22	AAH41019	PCR primer specifi
c	16	27	0.9	38	22	AAH41024	PCR primer specifi
c	17	27	0.9	38	22	AAH41033	PCR primer for met
c	18	27	0.9	38	22	AAH41034	PCR primer for met
c	19	25	0.9	610	24	ABQ44966	Oligonucleotide fo
c	20	25	0.9	610	24	ABQ44967	Oligonucleotide fo
c	21	23	0.8	2114	20	AAH19958	Rat rango-76 encod
c	22	22	0.8	1790	21	AAH48015	Zea mays DNA fragm
c	23	22	0.8	1813	21	AAH48015	Fragment of diamn
c	24	21	0.7	1827	21	AAH98952	Human pancreatic c
c	25	20	0.7	607	23	ABV48054	Human prostate exp
c	26	20	0.7	765	23	ABV18267	Human prostate exp
c	27	20	0.7	1244	21	AAH47449	Arabidopsis thalia
c	28	20	0.7	1247	21	AAH39225	Arabidopsis thalia
c	29	20	0.7	1802	23	ABLO1979	Drosophila melanog
c	30	20	0.7	3802	23	ABLO1979	Drosophila melanog
c	31	20	0.7	5880	22	AAH68687	Pseudomonas putida
c	32	20	0.7	11355	22	AAH68692	Pseudomonas putida
c	33	19	0.7	423	21	AAH95829	Bovine metalloprot
c	34	19	0.7	431	21	AAH95857	Bovine metalloprot
c	35	19	0.7	493	23	AAH86803	Human pancreatic c
c	36	19	0.7	518	23	AAH86803	DNA encoding novel
c	37	19	0.7	1046	22	AAH88168	CNS disorder-relat
c	38	19	0.7	1063	24	ABQ25996	Human protein HP10
c	39	19	0.7	1063	24	ABQ25996	Oligonucleotide fo
c	40	19	0.7	2089	24	ABQ25997	Oligonucleotide fo
c	41	19	0.7	2586	23	ABLO4141	Mouse ischaemic co
c	42	19	0.7	4479	24	ABN60018	Drosophila melanog
c	43	19	0.7	4533	21	AAA53920	Novel human coding
c	44	19	0.7	8139	22	ABA17510	Type III adenylyl
c	45	19	0.7	8139	22	ABA17511	Human nervous syst

ALIGNMENTS

RESULT 1	AAH41003	AAH41003 standard; cDNA; 2853 BP.
ID	AAH41003	standard; cDNA; 2853 BP.
XX	AC	AAH41003;
XX	DF	23-AUG-2001 (first entry)
XX	DE	Human metalloprotease MDTs6 cDNA.
XX	KW	Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;
XX	KW	osteopathic; antiarthritic; ss.
XX	OS	Homo sapiens.
XX	XX	Key
XX	XX	Location/Qualifiers
XX	XX	CDS
XX	XX	1..2853
XX	XX	/tag- a
XX	XX	/product= "Metalloprotease"
XX	XX	/note= "Aggrecanase activity"
XX	XX	WO200134785-A1.
XX	XX	17-MAY-2001.
XX	XX	10-NOV-2000; 2000WO-JP07917.
XX	XX	11-NOV-1999; 99JP-0321740.
XX	XX	16-MAY-2000; 2000JP-0144020.
XX	XX	(YAMA) YAMANOUCHI PHARM CO LTD.
XX	XX	(KAZU-) KAZUSA DNA RES INST.

XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
 XX WPI; 2001-343602/36.
 DR P-PSDB; AAG62299.
 XX Metalloprotease with aggrecanase activity for treating joint diseases
 PT especially osteoarthritis
 XX
 XX Example 2; Page 60-61; 85pp; Japanese.
 XX This invention relates to a metalloprotease with aggrecanase activity.
 CC The invention includes protein and DNA sequences of the metalloprotease,
 CC vectors containing the DNA, host cells transformed by the vectors, and
 CC antibodies directed against the metalloprotease. The antibodies, protein
 CC and DNA sequences can be used in the treatment and prevention of joint
 CC diseases, particularly osteoarthritis. The treatment may result in
 CC osteopathic and antiarthritic activity. The present sequence represents
 CC cDNA encoding the metalloprotease termed MDT56.
 XX
 SQ Sequence 2853 BP; 522 A; 920 C; 903 G; 508 T; 0 other;
 Query Match 100.0%; Score 2853; DB 22; Length 2853;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2853; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGCTTTTGTGGGATCTTAACCTGGCTTCGCGGGGAAACCGCTGGAGGCTTTAG 60
 DB 1 ATGCTTTTGTGGGATCTTAACCTGGCTTCGCGGGGAAACCGCTGGAGGCTTTAG 60
 QY 61 CCAGAGCGGAGGTAGTCTCCATCCGACTGGACCGCGGACATTAAACGCGCGGCTAC 120
 DB 61 CCAGAGCGGAGGTAGTCTCCATCCGACTGGACCGCGGACATTAAACGCGCGGCTAC 120
 QY 121 TACTGGCGGGTCCGAGGACTCCGGGGATCAGGGACTCATTTTCAGATCAGACATTT 180
 DB 121 TACTGGCGGGTCCGAGGACTCCGGGGATCAGGGACTCATTTTCAGATCAGACATTT 180
 QY 181 CAGGAGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCGCGCTTCTCC 240
 DB 181 CAGGAGACTTTTACCTACACCTGACCGCGGATGCTCAGTTCTTGGCTCCGCGCTTCTCC 240
 QY 241 ACTGAGCATCTGGCGTCCCGCTCCAGGGCTCACCGGGGCTTTCAGACCTTGCACGC 300
 DB 241 ACTGAGCATCTGGCGTCCCGCTCCAGGGCTCACCGGGGCTTTCAGACCTTGCACGC 300
 QY 301 TGCTTCTATTCTGGGACGTTGAACCGCGAGCTCGCTTCGCTGCTGTGAGCCTGTGC 360
 DB 301 TGCTTCTATTCTGGGACGTTGAACCGCGAGCTCGCTTCGCTGCTGTGAGCCTGTGC 360
 QY 361 GGGGGTCCCGGAGCCTTTGGCTACCGAGCGCGGAGTATGATATAGCCCGCTGCC 420
 DB 361 GGGGGTCCCGGAGCCTTTGGCTACCGAGCGCGGAGTATGATATAGCCCGCTGCC 420
 QY 421 AATGCTAGCGCGCGGCGGCGAGCAACAGCCAGCGGCGACACCTTCTCCAGCGCGG 480
 DB 421 AATGCTAGCGCGCGGCGGCGAGCAACAGCCAGCGGCGACACCTTCTCCAGCGCGG 480
 QY 481 GGTGTTCGGGCGGCGCTTCGGAGAGCCACCTCTCGCTCGCGGGTGGCTTCGGGCTGG 540
 DB 481 GGTGTTCGGGCGGCGCTTCGGAGAGCCACCTCTCGCTCGCGGGTGGCTTCGGGCTGG 540
 QY 541 AACCCCGCATCTACGGGCGCTTGAACCTTACAGCCGCGGCGGGGCTTCGGGGAG 600
 DB 541 AACCCCGCATCTACGGGCGCTTGAACCTTACAGCCGCGGCGGGGCTTCGGGGAG 600
 QY 601 AGTGTAGCGCGCGGCTTCGGGCGCGCAAGGTTTCGTATCCCGCGGTAGTG 660
 DB 601 AGTGTAGCGCGCGGCTTCGGGCGCGCAAGGTTTCGTATCCCGCGGTAGTG 660
 QY 661 GAGACGCTGTGTGGGAGGAGTCAATGATGTAAGTTCCACGGCGGCGGACCTTGGACAT 720
 DB 661 GAGACGCTGTGTGGGAGGAGTCAATGATGTAAGTTCCACGGCGGCGGACCTTGGACAT 720

QY 721 TATCTGCTGACGCTGCTGGCAACGCGCGGACTCTACCGCCATCCAGCATCCTCAAC 780
 DB 721 TATCTGCTGACGCTGCTGGCAACGCGCGGACTCTACCGCCATCCAGCATCCTCAAC 780
 QY 781 CCCATCAACATCGTGTGGTCAAGGTGCTTCTTAGAGATCGTACTCGGGGCCAAG 840
 DB 781 CCCATCAACATCGTGTGGTCAAGGTGCTTCTTAGAGATCGTACTCGGGGCCAAG 840
 QY 841 GTCACCGCAATGCGGCGCTGACGCTGGCAACTTCTGTGCTGGCAGAGAACTGAAC 900
 DB 841 GTCACCGCAATGCGGCGCTGACGCTGGCAACTTCTGTGCTGGCAGAGAACTGAAC 900
 QY 901 AAAGTGTAGTGAACAGCACCCGAGTACTTGGACACTTGCATCTCTTCCACAGGAGAC 960
 DB 901 AAAGTGTAGTGAACAGCACCCGAGTACTTGGACACTTGCATCTCTTCCACAGGAGAC 960
 QY 961 CTGTGTGAGGCGACCACTGTGACACCTGGCATGGCTGATGTGGTACTATCTGTGAC 1020
 DB 961 CTGTGTGAGGCGACCACTGTGACACCTGGCATGGCTGATGTGGTACTATCTGTGAC 1020
 QY 1021 CCCAAGAGAACTGCTCTGTCTATTGAGGACGATGGGCTTCCATCAGCCTTCCACTGCC 1080
 DB 1021 CCCAAGAGAACTGCTCTGTCTATTGAGGACGATGGGCTTCCATCAGCCTTCCACTGCC 1080
 QY 1081 CACGAGCTGGGCGACGTTTCAACATGCCCATGACAAATGTGAAAGTCTGTGAGGAGTG 1140
 DB 1081 CACGAGCTGGGCGACGTTTCAACATGCCCATGACAAATGTGAAAGTCTGTGAGGAGTG 1140
 QY 1141 TTTGGAGAGCTCCGAGCAACACATGATGTCCCGACCTTCATCCAGATCGACCGTGCC 1200
 DB 1141 TTTGGAGAGCTCCGAGCAACACATGATGTCCCGACCTTCATCCAGATCGACCGTGCC 1200
 QY 1201 AACCCCTGCTCAGCTGCGAGTGTCCCATCATCAGCAGTTCCTGGACAGCGGACGGT 1260
 DB 1201 AACCCCTGCTCAGCTGCGAGTGTCCCATCATCAGCAGTTCCTGGACAGCGGACGGT 1260
 QY 1261 GACTGCTCTGGGCAACCCAGAGCCCATCTCCCTGCGCGGAGGATCTGCCGGGCGCC 1320
 DB 1261 GACTGCTCTGGGCAACCCAGAGCCCATCTCCCTGCGCGGAGGATCTGCCGGGCGCC 1320
 QY 1321 AGCTACACCTCAGCAGCAGTGTGGCTTTTGGCGTGGGCTCCAAGCTCTGTCT 1380
 DB 1321 AGCTACACCTCAGCAGCAGTGTGGCTTTTGGCGTGGGCTCCAAGCTCTGTCT 1380
 QY 1381 TACATGCACTGTGACCAAGCTGTGTGACCGGAGAGCCAGGACACATGCTGTGTC 1440
 DB 1381 TACATGCACTGTGACCAAGCTGTGTGACCGGAGAGCCAGGACACATGCTGTGTC 1440
 QY 1441 CAGACCGCGCTTCCCTGGGCGGATGGCAGCTGTGGGAGGCAAGCTCTGCCCTC 1500
 DB 1441 CAGACCGCGCTTCCCTGGGCGGATGGCAGCTGTGGGAGGCAAGCTCTGCCCTC 1500
 QY 1501 AAAGGGGCTGCTGGAGAGACACACCTCAACAGCAGAGGCTGGATGGTTCCTGGGCC 1560
 DB 1501 AAAGGGGCTGCTGGAGAGACACACCTCAACAGCAGAGGCTGGATGGTTCCTGGGCC 1560
 QY 1561 AAATGGGATCCCTATGGCCCTGCTGCGGCATGTGTGGGGGCTGCACTCGCCAGG 1620
 DB 1561 AAATGGGATCCCTATGGCCCTGCTGCGGCATGTGTGGGGGCTGCACTCGCCAGG 1620
 QY 1621 AGGAGTGCACCAACCCACCTGCGGAGGCTGCTGCGGAGGAGTACTGAGGGAGT 1680
 DB 1621 AGGAGTGCACCAACCCACCTGCGGAGGCTGCTGCGGAGGAGTACTGAGGGAGT 1680
 QY 1681 AAATACCGATCTGCAACCTGGAGCCCTGCCCGAGCTCAGCCTCGGAAAGAGCTTCGGG 1740
 DB 1681 AAATACCGATCTGCAACCTGGAGCCCTGCCCGAGCTCAGCCTCGGAAAGAGCTTCGGG 1740
 QY 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACACCGAGTACACCGAGTCTACTCTGCC 1800
 DB 1741 GAGGAGCAGTGTGAGGCTTTCAACGGCTACACCGAGTACACCGAGTCTACTCTGCC 1800

QY 1801 GTGCATGGGTGCCCAAGTACTCCGGCTGTCTCCCGGGACAAGTGAAGCTCATCTGC 1860
Db 1801 GTGGCATGGGTGCCCAAGTACTCCGGCTGTCTCCCGGGACAAGTGAAGCTCATCTGC 1860
QY 1861 CGAGCAATGGCACTGGCTACTTCTATGTGCTGGGACCCCAAGGTGGTGGACGGCAGGTG 1920
Db 1861 CGAGCAATGGCACTGGCTACTTCTATGTGCTGGGACCCCAAGGTGGTGGACGGCAGGTG 1920
QY 1921 TGCTCTCTCTGACTCCACCTCCGCTCTGTGCTCAAGCAAGTGCATCAAGCTGGCTGTAT 1980
Db 1921 TGCTCTCTCTGACTCCACCTCCGCTCTGTGCTCAAGCAAGTGCATCAAGCTGGCTGTAT 1980
QY 1981 GGAACCTGGGCTTCAAGAAAGATTCGACAAAGTGTGGGTGTGGGGAGACAAATAG 2040
Db 1981 GGAACCTGGGCTTCAAGAAAGATTCGACAAAGTGTGGGTGTGGGGAGACAAATAG 2040
QY 2041 AGCTGCAAGAGTGTACTGACTCTTCAAGCAAGTGCATCAAGCTGGCTGTAT 2100
Db 2041 AGCTGCAAGAGTGTACTGACTCTTCAAGCAAGTGCATCAAGCTGGCTGTAT 2100
QY 2101 GCATCCCCCGAGGGCCCTCAAGCATCGACATCCGCCAGCGGTTACAAAGGGCTGATC 2160
Db 2101 GCATCCCCCGAGGGCCCTCAAGCATCGACATCCGCCAGCGGTTACAAAGGGCTGATC 2160
QY 2161 GGGATGACAACTACTTGGCTCTGAAGACACGCCAAGCAAGTACTCTCAAGGGCAT 2220
Db 2161 GGGATGACAACTACTTGGCTCTGAAGACACGCCAAGCAAGTACTCTCAAGGGCAT 2220
QY 2221 TTCGTGGTGTGCGGTGAGCGGACCTGCTGAGGCGAGTCTGCTCGGTACAGC 2280
Db 2221 TTCGTGGTGTGCGGTGAGCGGACCTGCTGAGGCGAGTCTGCTCGGTACAGC 2280
QY 2281 GGCAGGGCACAGCGGTGAGAGCTGCAAGGCTTCCGGGCCATCTCGAGCCGCTGACC 2340
Db 2281 GGCAGGGCACAGCGGTGAGAGCTGCAAGGCTTCCGGGCCATCTCGAGCCGCTGACC 2340
QY 2341 GTGAGGTCTCTCGTGGGGAAGATGACCGCCCGGGTCCGCTACTCTCTTACTG 2400
Db 2341 GTGAGGTCTCTCGTGGGGAAGATGACCGCCCGGGTCCGCTACTCTCTTACTG 2400
QY 2401 CCCAAGAGCTCGGGAGGACAAGTCTCTCATCCCAAGGACCCCGGGACCTCTCTC 2460
Db 2401 CCCAAGAGCTCGGGAGGACAAGTCTCTCATCCCAAGGACCCCGGGACCTCTCTC 2460
QY 2461 TTGCACACAGCTCTCAGCTCTCAACCAAGTGGAGCAGCGGAGCAGAGCCGCT 2520
Db 2461 TTGCACACAGCTCTCAGCTCTCAACCAAGTGGAGCAGCGGAGCAGAGCCGCT 2520
QY 2521 GCACGCTGGGTGGCTGGCAGCTGGGGGCGCTGCTCCGCGAGCTGCGGAGTGGCTGAG 2580
Db 2521 GCACGCTGGGTGGCTGGCAGCTGGGGGCGCTGCTCCGCGAGCTGCGGAGTGGCTGAG 2580
QY 2581 AAGCGGGGCTGGAGTCCGGGGTTCGCCGGGAGCAGCGTCCCTGCTGTGATGCA 2640
Db 2581 AAGCGGGGCTGGAGTCCGGGGTTCGCCGGGAGCAGCGTCCCTGCTGTGATGCA 2640
QY 2641 GCCATCGCCGCTGGAGACACAGCTTCCGGGGAGCCTGCCCGAGCCTGGGAGCTCAGC 2700
Db 2641 GCCATCGCCGCTGGAGACACAGCTTCCGGGGAGCCTGCCCGAGCCTGGGAGCTCAGC 2700
QY 2701 GCCTGGTCAACCTGCTCCAAAGAGTTCGGGGGAGTTCAGAGGCGCTCACTCAAGTGT 2760
Db 2701 GCCTGGTCAACCTGCTCCAAAGAGTTCGGGGGAGTTCAGAGGCGCTCACTCAAGTGT 2760
QY 2761 GTGGCCACAGGAGCGGCTGCTGGCCCGGGACCAAGTGCACCTTGCACCGGAGCCGAG 2820
Db 2761 GTGGCCACAGGAGCGGCTGCTGGCCCGGGACCAAGTGCACCTTGCACCGGAGCCGAG 2820
QY 2821 GAGCTGGACTTCTGGTCTTCTAGGCGCTGTGTA 2853
Db 2821 GAGCTGGACTTCTGGTCTTCTAGGCGCTGTGTA 2853

RESULT 2
AAD35569
ID AAD35569 standard; cDNA; 2853 BP.
XX
XX AAD35569;
XX AC
XX
XX
DT 26-JUL-2002 (first entry)
XX
XX Human protease cDNA #2.
XX
XX Human; novel human protein; NHP; protease; biological disorder; obesity;
XX high blood pressure; arthritis; connective tissue disorder; infertility;
XX gene therapy; enzyme; gene; ss.
XX
XX Homo sapiens.
XX
XX
FH Key Location/Qualifiers
CDS 1..2853
FT FT
FT /*tag= a
FT /product= "Human protease #2"
XX
XX WO200226949-A2.
XX
XX 04-APR-2002.
XX
XX 27-SEP-2001; 2001WO-US30350.
XX
XX 29-SEP-2000; 2000US-236689P.
XX
XX (LEXI-) LEXICON GENETICS INC.
XX
XX
XX Friddle CJ, Hilbun E;
XX
XX WPI; 2002-372123/40.
XX
XX P-PSDB; AAD22541.
XX
XX Novel nucleic acid encoding a human protease, useful as a hybridization
XX probe for screening libraries and assessing gene expression patterns -
XX
XX Claim 1; Page 35-36; 41pp; English.
XX
XX The present sequence is a cDNA encoding novel human protein (NHP),
XX human protease. NHPs share structural similarity with animal proteases
XX particularly zinc metalloproteases. Sequences of the invention are
XX useful in therapeutic, diagnostic and pharmacogenomic applications.
XX NHP polynucleotides are used as hybridisation probes for screening
XX libraries and assessing gene expression patterns. They can also be
XX used for treating related biological disorders such as obesity, high
XX blood pressure, arthritis, connective tissue disorders and infertility.
XX They are also used in gene therapy.
XX
XX Sequence 2853 BP; 522 A; 920 C; 903 G; 508 T; 0 other;
Query Match 94.5%; Score 2695; DB 24; Length 2853;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2795; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 57 TGAGCCAGAGCGGGAGTAGTCTCCATCCGACTGGACCGGACATTAAACGGCGCG 116
Db 57 TGAGCCAGAGCGGGAGTAGTCTCCATCCGACTGGACCGGACATTAAACGGCGCG 116
QY 117 CTACTACTGGCGGGTCCCGAGGACTCCCGGGATCAGGACTCATTTTCATGATACAGC 176
Db 117 CTACTACTGGCGGGTCCCGAGGACTCCCGGGATCAGGACTCATTTTCATGATACAGC 176
QY 177 ATTTTACAGGAGACTTTTACTACACTGACCGCGGATGCTCAGTCTTGGTTCGGCTT 236
Db 177 ATTTTACAGGAGACTTTTACTACACTGACCGCGGATGCTCAGTCTTGGTTCGGCTT 236
QY 237 CTCCTACTGAGCATCTGGGGTTCCTCCCTCCAGGGGCTCAGCGGGGCTCTTACAGCTGG 296
Db 237 CTCCTACTGAGCATCTGGGGTTCCTCCCTCCAGGGGCTCAGCGGGGCTCTTACAGCTGG 296

Db 2457 TGTCTGCAACAGCGTCTCTAGCCCTCTCCAAACCAGGTGGAGCAGCGGACGACAGGC 2516
 QY 2517 CCCTGACGCTGGGTGCTGCGACGTGGGGCGGTGCTCCGCGAGCTGGGCGAGTGGCCT 2576
 Db 2517 CCCTGACGCTGGGTGCTGCGACGTGGGGCGGTGCTCCGCGAGTGGGCGAGTGGCCT 2576
 QY 2577 GCAGAACGGCGGTGACTGCGGGGCTCCGCGGCGACGCGACGCTCCCTGCTGTGA 2636
 Db 2577 GCAGAACGGCGGTGACTGCTGGGGCTCCGCGGCGACGCGACGCTCCCTGCTGTGA 2636
 QY 2637 TGACGCCCATCGGCCGTGAGACACAAAGCTCGGGGAGCCCTGCCACCTGGGAGCT 2696
 Db 2637 TGACGCCCATCGGCCGTGAGACACAAAGCTCGGGGAGCCCTGCCACCTGGGAGCT 2696
 QY 2697 CAGCGCTGCTACCTGCTCCAAAGAGCTCGGCGGGGATTTTCAGAGCGCTCACTGAA 2756
 Db 2697 CAGCGCTGCTACCTGCTCCAAAGAGCTCGGCGGGGATTTTCAGAGCGCTCACTGAA 2756
 QY 2757 GTGTGTGGCCACGGAGCGCGGTGCTGGCCGGGACCAAGTTCACCGCAAGCC 2816
 Db 2757 GTGTGTGGCCACGGAGCGCGGTGCTGGCCGGGACCAAGTTCACCGCAAGCC 2816
 QY 2817 CCAGGAGCTGACTTCTGCGCTCTGAGCGGTGCTGA 2853
 Db 2817 CCAGGAGCTGACTTCTGCGCTCTGAGCGGTGCTGA 2853

RESULT 3
 AAD35571

ID AAD35571 standard; cDNA; 3446 BP.

AC AAD35571;

XX 26-JUL-2002 (first entry)

DE Human protease cDNA #4.

XX Human; novel human protein; NHP; protease; biological disorder; obesity;
 KW high blood pressure; arthritis; connective tissue disorder; infertility;
 KW gene therapy; enzyme; ss.

OS Homo sapiens.

PN WO200226949-A2.

XX 04-APR-2002.

PF 27-SEP-2001; 2001WO-US30350.

XX 29-SEP-2000; 2000US-236689P.

PR (LEXI-) LEXICON GENETICS INC.

PA Friddle CJ, Hilbun E;

XX WPI; 2002-372123/40.

PT Novel nucleic acid encoding a human protease, useful as a hybridization
 probe for screening libraries and assessing gene expression patterns -

XX Disclosure; Page 40-41; 41pp; English.

XX The present sequence is a cDNA encoding novel human protein (NHP),
 CC human protease. NHPs share structural similarity with animal proteases
 CC particularly zinc metalloproteases. Sequences of the invention are
 CC useful in therapeutic, diagnostic and pharmacogenomic applications.
 CC NHP polynucleotides are used as hybridization probes for screening
 CC libraries and assessing gene expression patterns. They can also be
 CC used for treating related biological disorders such as obesity, high
 CC blood pressure, arthritis, connective tissue disorders and infertility.
 CC They are also used in gene therapy.

XX Sequence 3446 BP; 612 A; 1114 C; 1101 G; 619 T; 0 other;

Query Match 94.5%; Score 2695; DB 24; Length 3446;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2795; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 57 TCAGCCACAGCGGGAGGTAGTCTCCCATCCGACTGACCGCGGACATTAACGGCCGCG 116
 Db 453 TCAGCCACAGCGGGAGGTAGTCTCCCATCCGACTGACCGCGGACATTAACGGCCGCG 512
 QY 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCAATTTTTCAGATCACAGC 176
 Db 513 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGCTCAATTTTTCAGATCACAGC 572
 QY 177 ATTTTCAGGAGGACTTTTACCTACACCTACGCGCGGATCTCAGTCTCTGGCTCCGCGCT 236
 Db 573 ATTTTCAGGAGGACTTTTACCTACACCTACGCGCGGATCTCAGTCTCTGGCTCCGCGCT 632
 QY 237 CTCCTACTGAGCATCTGGCGCTCCCTCCAGGGGCTACCGGGGCTCTTCAGACCTGCG 296
 Db 633 CTCCTACTGAGCATCTGGCGCTCCCTCCAGGGGCTACCGGGGCTCTTCAGACCTGCG 692
 QY 297 AGCTGCTTCTATTCTGGGAGCTGAACCGCGAGCGGACTCTGTTGGTGTGTGTAGCCT 356
 Db 693 AGCTGCTTCTATTCTGGGAGCTGAACCGCGAGCGGACTCTGTTGGTGTGTGTAGCCT 752
 QY 357 GTGCGGGGGCTCCGCGAGGCTTTTGGCTACCGGGGCTACCGGGGCTCTTCAGACCTGCG 416
 Db 753 GTGCGGGGGCTCCGCGAGGCTTTTGGCTACCGGGGCTACCGGGGCTCTTCAGACCTGCG 812
 QY 417 GCCCAATGCTAGCGCGCGCGGCGGCGACGCAACAGCGCGGCGCACACCTTCTTCAGCG 476
 Db 813 GCCCAATGCTAGCGCGCGCGGCGGCGACGCAACAGCGCGGCGCACACCTTCTTCAGCG 872
 QY 477 CCGGGGTCTTCGGGCGGGCTTCGGGAGACCCACCTCTCGCTTCGGGTGCGGCTCGG 536
 Db 873 CCGGGGTCTTCGGGCGGGCTTCGGGAGACCCACCTCTCGCTTCGGGTGCGGCTCGG 932
 QY 537 CTGGAACCCCGCATCTTACGCGCGCTTACAGCGCGGCGGCGGCGGCTTCGCG 596
 Db 933 CTGGAACCCCGCATCTTACGCGCGCTTACAGCGCGGCGGCGGCGGCTTCGCG 992
 QY 597 GGAGAGTCGTAGCGCGCGAGTCTGGGCGCGCAAGCGTTTCGTCTATCCCGCGGTA 656
 Db 993 GGAGAGTCGTAGCGCGCGAGTCTGGGCGCGCAAGCGTTTCGTCTATCCCGCGGTA 1052
 QY 657 CGTGAGAGCTGTGTGCTGGGAGGAGTCAATGTTCAAGTTCACGCGCGGCGGAGCTGGA 716
 Db 1053 CGTGAGAGCTGTGTGCTGGGAGGAGTCAATGTTCAAGTTCACGCGCGGCGGAGCTGGA 1112
 QY 717 ACATTATCTGTGACGCTGTGGCAACGCGCGGCGGAGTCTTACCGCATCCAGCATCCT 776
 Db 1113 ACATTATCTGTGACGCTGTGGCAACGCGCGGAGTCTTACCGCATCCAGCATCCT 1172
 QY 777 CAACCCCATCAACATCGTTGGTCAAGGTGCTCTTTTAGAGATCGTGACTCCGGGCG 836
 Db 1173 CAACCCCATCAACATCGTTGGTCAAGGTGCTCTTTTAGAGATCGTGACTCCGGGCG 1232
 QY 837 CAAGGTACCGGCAATCGGCGCTGACGCTGCGCAACTTCTGTGCTTGGCGAGAGAGCT 896
 Db 1233 CAAGGTACCGGCAATCGGCGCTGACGCTGCGCAACTTCTGTGCTTGGCGAGAGAGCT 1292
 QY 897 GAACAAAGTGTAGTGACAAGCACCCGAGTACTGGGACACTGCCATCTCTTTCACGAGCA 956
 Db 1293 GAACAAAGTGTAGTGACAAGCACCCGAGTACTGGGACACTGCCATCTCTTTCACGAGCA 1352
 QY 957 GGACCTGTGGAGGCCACCACTGTGACACCCCTGGGATGGCTGATGGGTACATGTG 1016
 Db 1353 GGACCTGTGGAGGCCACCACTGTGACACCCCTGGGATGGCTGATGGGTACATGTG 1412
 QY 1017 TGACCCCAAGAGAGCTGCTCTGTCATTGAGGAGGATGGGCTTCCATCAGCCTTACCAC 1076
 Db 1413 TGACCCCAAGAGAGCTGCTCTGTCATTGAGGAGGATGGGCTTCCATCAGCCTTACCAC 1472

```
QY 1077 TGCCACGAGCTGGGCCAGCTGTTCACATGCCCATGACAAATGTGAAGTCTGTGAGGA 1136
|
|
|
Db 1473 TGCCACGAGCTGGGCCAGCTGTTCACATGCCCATGACAAATGTGAAGTCTGTGAGGA 1532
|
|
|
QY 1137 GGTGTTTGGGAAGCTCCGAGCAACACATGATGTCCCGAGCCCTCATCCAGATCGACGG 1196
|
|
|
Db 1533 GGTGTTTGGGAAGCTCCGAGCAACACATGATGTCCCGAGCCCTCATCCAGATCGACGG 1592
|
|
|
QY 1197 TGCCAAACCCCTGGTCAAGCTGCAGTGTGCCATCATCACCGACTTCCCTGGACAGCGGCA 1256
|
|
|
Db 1593 TGCCAAACCCCTGGTCAAGCTGCAGTGTGCCATCATCACCGACTTCCCTGGACAGCGGCA 1652
|
|
|
QY 1257 CGGTGACTGCTCTTGAGCAACCCAGCAAGCCATCTCCCTGCCCGAGGATCTCCCGGG 1316
|
|
|
Db 1653 CGGTGACTGCTCTTGAGCAACCCAGCAAGCCATCTCCCTGCCCGAGGATCTCCCGGG 1712
|
|
|
QY 1317 CGCCAGCTACACCTGAGCCAGAGTGCAGCTGGTGTGGCTGGGCTCCAAAGCCCTG 1376
|
|
|
Db 1713 CGCCAGCTACACCTGAGCCAGAGTGCAGCTGGTGTGGCTGGGCTCCAAAGCCCTG 1772
|
|
|
QY 1377 TCCTTACATGCACTACTGCACCAAGCTGTGTGCACCGGAAGGCCAAGGGACAGATGGT 1436
|
|
|
Db 1773 TCCTTACATGCACTACTGCACCAAGCTGTGTGCACCGGAAGGCCAAGGGACAGATGGT 1832
|
|
|
QY 1437 GTCCAGACCCGCCACTTCCCTGGGCCGATGGCAGCAGTGTGGCGAGGGCAAGCTGTG 1496
|
|
|
Db 1833 GTCCAGACCCGCCACTTCCCTGGGCCGATGGCAGCAGTGTGGCGAGGGCAAGCTGTG 1892
|
|
|
QY 1497 CCTCAAGGGGCTGGTGGAGAGACACACCTCAACAGCAGACAGGTGGATGTTCTGTG 1556
|
|
|
Db 1893 CCTCAAGGGGCTGGTGGAGAGACACACCTCAACAGCAGACAGGTGGATGTTCTGTG 1952
|
|
|
QY 1557 GGCCAAATGGATCCCTATGCCCCCTGCTCGGCACATGTGTGGGGGCGTGCAGTGGC 1616
|
|
|
Db 1953 GGCCAAATGGATCCCTATGCCCCCTGCTCGGCACATGTGTGGGGGCGTGCAGTGGC 2012
|
|
|
QY 1617 CAGAGGAGTGCACCAACCCACCCCTGCCAAGCGGGGCAAGTACTGCGAGGGAGTGTAG 1676
|
|
|
Db 2013 CAGAGGAGTGCACCAACCCACCCCTGCCAAGCGGGGCAAGTACTGCGAGGGAGTGTAG 2072
|
|
|
QY 1677 GGTGAATACCGATCCTGCAACCTGGAGCCCTGCCCGAGCTCAGCCCTCCGGAAAGAGCTT 1736
|
|
|
Db 2073 GGTGAATACCGATCCTGCAACCTGGAGCCCTGCCCGAGCTCAGCCCTCCGGAAAGAGCTT 2132
|
|
|
QY 1737 CCGGGAGGAGTGTGAGGCTTTTCAAGGCTTACAACGAGTACAACAGCACCACCCGCTCACTCT 1796
|
|
|
Db 2133 CCGGGAGGAGTGTGAGGCTTTTCAAGGCTTACAACGAGTACAACAGCACCACCCGCTCACTCT 2192
|
|
|
QY 1797 CGCCGTGGCATGGTGGCCCAAGTACTCCGCGGTGTCTCCCGGCAAGTGAAGCTCAT 1856
|
|
|
Db 2193 CGCCGTGGCATGGTGGCCCAAGTACTCCGCGGTGTCTCCCGGCAAGTGAAGCTCAT 2252
|
|
|
QY 1857 CTGCGAGGCAATGGCACTGGCTACTTCTATGTGTGCTGGCACCAAGGTGTGGACGGCAC 1916
|
|
|
Db 2253 CTGCGAGGCAATGGCACTGGCTACTTCTATGTGTGCTGGCACCAAGGTGTGGACGGCAC 2312
|
|
|
QY 1917 GCTGTGCTCTCTGACTCCACCTCCGCTGTGTGTCTTCCAAAGGCAAGTGCATCAAGGCTGGCTG 1976
|
|
|
Db 2313 GCTGTGCTCTCTGACTCCACCTCCGCTGTGTGTCTTCCAAAGGCAAGTGCATCAAGGCTGGCTG 2372
|
|
|
QY 1977 TGATGGAACTGGGCTCCAAAGAGAGATTCGACAGTGTGGGCTGTGGGGGAGACAA 2036
|
|
|
Db 2373 TGATGGAACTGGGCTCCAAAGAGAGATTCGACAGTGTGGGCTGTGGGGGAGACAA 2432
|
|
|
QY 2037 TAAGAGCTGCAAGAGTGTGACTGACTTCTTCAACAGCCCATCATGGCTTACAAATTCGT 2096
|
|
|
Db 2433 TAAGAGCTGCAAGAGTGTGACTGACTTCTTCAACAGCCCATCATGGCTTACAAATTCGT 2492
|
|
|
QY 2097 GGTGGCCATCCCGCAGGCGCCCTCAAGCATTCGACATCCCGCAGCGGTTTACAAAGGCT 2156
|
|
|
Db 2493 GGTGGCCATCCCGCAGGCGCCCTCAAGCATTCGACATCCCGCAGCGGTTTACAAAGGCT 2552
|
|
|
QY 2157 GATCGGGATGACAACTACCTGGCTCTGAGAACAGCCCAAGGCAAGTACCTGTCTCAACGG 2216
|
|
|
```

```
Db 2553 GATCGGGATGACAACTACCTGGCTCTGAAGAACAGCAAGGCAAGTACCTGTCTCAACGG 2612
|
|
|
QY 2217 GCATTTCTGTGTCTCGCGGTGGAGCGGACCTGGTGGTGAAGGGAGTGTCTGTGCGGTA 2276
|
|
|
Db 2613 GCATTTCTGTGTCTCGCGGTGGAGCGGACCTGGTGGTGAAGGGAGTGTCTGTGCGGTA 2672
|
|
|
QY 2277 CAGCGGACCGGGACAGCGGTGGAGAGCCTGCAGGCTTCCCGGCCCATCTGTGAGCGGCT 2336
|
|
|
Db 2673 CAGCGGACCGGGACAGCGGTGGAGAGCCTGCAGGCTTCCCGGCCCATCTGTGAGCGGCT 2732
|
|
|
QY 2337 GACCGTGGAGTCTCTCCGTGGGGAAGATGACACCCGCCCGGGTCCGCTACTCTTCTA 2396
|
|
|
Db 2733 GACCGTGGAGTCTCTCCGTGGGGAAGATGACACCCGCCCGGGTCCGCTACTCTTCTA 2792
|
|
|
QY 2397 TCTGCCAAAGAGCTCGGGAGGACAACTCTCTCATCCCAAGGACCCCGGGGACCCCTC 2456
|
|
|
Db 2793 TCTGCCAAAGAGCTCGGGAGGACAACTCTCTCATCCCAAGGACCCCGGGGACCCCTC 2852
|
|
|
QY 2457 TGTCTTGACAAACAGCGTCTCAGCCTCTCCAAACAGGTGGAGCAGCGGACGACAGGCC 2516
|
|
|
Db 2853 TGTCTTGACAAACAGCGTCTCAGCCTCTCCAAACAGGTGGAGCAGCGGACGACAGGCC 2912
|
|
|
QY 2517 CCCTGACAGCTGGTGGCTGGCAGCTGGGGCCGTCTCCGAGAGCTGCGGAGTGGCCCT 2576
|
|
|
Db 2913 CCCTGACAGCTGGTGGCTGGCAGCTGGGGCCGTCTCCGAGAGCTGCGGAGTGGCCCT 2972
|
|
|
QY 2577 GCAGAAGCGGCGGTGGACTGCGCGGGGCTCCGCGGGCAGCGACGGTCCCTTGCTGTGA 2636
|
|
|
Db 2973 GCAGAAGCGGCGGTGGACTGCTCGGGGCTCCGCGGGCAGCGACGGTCCCTTGCTGA 3032
|
|
|
QY 2637 TGCAGCCATCGGCCGTGGAGACACAAAGCTGCGGGAGCCCTGCCCCACCTGGGAGCT 2696
|
|
|
Db 3033 TGCAGCCATCGGCCGTGGAGACACAAAGCTGCGGGAGCCCTGCCCCACCTGGGAGCT 3092
|
|
|
QY 2697 CAGGCGCTGGTCACTCTCTCAAGAGCTGCGGCGGGGATTTTCAGAGGGGCTCACTCAA 2756
|
|
|
Db 3093 CAGGCGCTGGTCACTCTCTCAAGAGCTGCGGCGGGGATTTTCAGAGGGGCTCACTCAA 3152
|
|
|
QY 2757 GTGTGTGGCCACGAGGCGCGGCTGTGGCCCGGGACAGTGCACACTTGCACCCCAAGCC 2816
|
|
|
Db 3153 GTGTGTGGCCACGAGGCGCGGCTGTGGCCCGGGACAGTGCACACTTGCACCCCAAGCC 3212
|
|
|
QY 2817 CCAGAGCTGGACTTCTCGCTCTGAGGCGGCTGTGA 2853
|
|
|
Db 3213 CCAGAGCTGGACTTCTCGCTCTGAGGCGGCTGTGA 3249
|
|
|
RESULT 4
ABK12894
ID ABK12894 standard; cdna; 2930 BP.
XX
AC ABK12894;
XX
DT 09-APR-2002 (first entry)
XX
DE Human protease PRS-11 cDNA sequence.
XX
KW Human; protease; PRS; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 75..2930
FT /*tag= a
FT /partial
FT /product= "Human protease PRS-11"
FT /note= "This sequence lacks a stop codon"
```


XX PN WO200198468-A2.
XX PD 27-DEC-2001.
XX PF 13-JUN-2001; 2001WO-US19178.
XX PR 16-JUN-2000; 2000US-2123336P.
XX PR 22-JUN-2000; 2000US-213955P.
XX PR 29-JUN-2000; 2000US-215396P.
XX PR 07-JUL-2000; 2000US-216821P.
XX PR 14-JUL-2000; 2000US-218946P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
PI Delegrane AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Wallia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
XX PI Kallick DA;
XX DR WPI: 2002-090437/12.
XX DR P-PSDB: AAU74751.
XX PT Twenty one human proteases (referred to as PRPS-1 to PRPS-21), useful
XX PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
XX PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
XX PT proliferative (e.g. cancer) disorders -
XX PS Claim 5; Page 168; 177pp; English.
XX CC The present invention relates to twenty one new human proteases,
XX CC referred to as PRPS-1 to PRPS-21. The PRPS polynucleotides and
XX CC polypeptides of the invention are useful in the diagnosis, treatment and
XX CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
XX CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
XX CC myocardial infarction, autoimmune/inflammatory e.g. acquired
XX CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
XX CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
XX CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
XX CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
XX CC endometriosis disorders. Numerous other examples of each disorder are
XX CC given in the specification. The present nucleic acid sequence encodes
XX CC the human protease PRPS-11 protein of the invention.
XX SQ Sequence 2930 BP; 529 A; 958 C; 931 G; 512 T; 0 other;

Query Match 73.2%; Score 2089; DB 24; Length 2930;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2379; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 57 TGAGCCAGAGCGGGAGGTAGTCTCCATCCGACTGGACCCGGACATTAAAGCGCGCG 116
DB 131 TGAGCCAGAGCGGGAGGTAGTCTCCATCCGACTGGACCCGGACATTAAAGCGCGCG 190
QY 117 CTACTACTGCGGGGTCGCGAGACTCCGGGATCAGGACATCTTTTCAGATCAGAC 176
DB 191 CTACTACTGCGGGGTCGCGAGACTCCGGGATCAGGACATCTTTTCAGATCAGAC 250
QY 177 ATTTCAGGAGGACTTTTACCTACACTGACCGCGGATGCTGATGCTGCGCTCCCGCTT 236
DB 251 ATTTCAGGAGGACTTTTACCTACACTGACCGCGGATGCTGATGCTGCGCTCCCGCTT 310
QY 237 CTCCACTGAGCATCTGGGCGTCCCGCTCCAGGGGTCACCGGGGCTCTTCAGACCTGCG 296
DB 311 CTCCACTGAGCATCTGGGCGTCCCGCTCCAGGGGTCACCGGGGCTCTTCAGACCTGCG 370
QY 297 AGCTCTCTTCTATCTGGGAGCTGACCCCGAGCGGACTCGTCTGCTGCTGAGCCT 356
DB 371 AGCTCTCTTCTATCTGGGAGCTGAAACCGGAGCGGACTCGTCTGCTGCTGAGCCT 430
QY 357 GTGCGGGGGCTCCCGAGGCTTTGGGTACCGAGGCGCGAGTAGTAGTATGCTAGCCGCT 416
|||||

DB 431 GTGCGGGGGCTCCCGCGAGCCTTTGGGTACCCAGGCGCGAGTAGTATGCTATTAGCCCGCT 490
QY 417 GCCAATAGTACGCGCGCGCGCGCAGCCCAACAGCCAGGCGCACACTTCTTCTCCAGCG 476
DB 491 GCCAATAGTACGCGCGCGCGCGCAGCCCAACAGCCAGGCGCACACTTCTTCTCCAGCG 550
QY 477 CGGGGTGTTCCGGGGGGGCTTCCGGAGAGCCACCTCTCGCTCGGGGTGGCTCGGG 536
DB 551 CGGGGTGTTCCGGGGGGGCTTCCGGAGAGCCACCTCTCGCTCGGGGGTGGCTCGGG 610
QY 537 CTGGAACCCCGCATCTCTACGGGCGCTGAGACCTTACAGCGCGCGGGGGTTCGG 596
DB 611 CTGGAACCCCGCATCTCTACGGGCGCTGAGACCTTACAGCGCGCGGGGGTTCGG 670
QY 597 GGAGAGTCTAGCCGGCGCAGGTCTGGGCGCGCCAAAGGCTTTCGTGTCTATCCCCCGGTA 656
DB 671 GGAGAGTCTAGCCGGCGCAGGTCTGGGCGCGCCAAAGGCTTTCGTGTCTATCCCCCGGTA 730
QY 657 CGTGAGAGCGCTGGTGGTCCGCGAGAGTCAATGTCAAGTTCACAGCGCGGAGCCTGGA 716
DB 731 CGTGAGAGCGCTGGTGGTCCGCGAGAGTCAATGTCAAGTTCACAGCGCGGAGCCTGGA 790
QY 717 ACATTATCTGTGACGCTGCTGCGCAACGGCGCGGACTCTACCGCCATCCAGCATCCT 776
DB 791 ACATTATCTGTGACGCTGCTGCGCAACGGCGCGGACTCTACCGCCATCCAGCATCCT 850
QY 777 CAACCCCATCAACATCGTTGGTCAAGTGTCTTTAGAGATCGTGACTCGGGGCC 836
DB 851 CAACCCCATCAACATCGTTGGTCAAGTGTCTTTAGAGATCGTGACTCGGGGCC 910
QY 837 CAAGTCTACCGGCAATCGGCGCTGACGCTGCGCAACGGCGCGGACTCTGCTGGCAGAGAACT 896
DB 911 CAAGTCTACCGGCAATCGGCGCTGACGCTGCGCAACGGCGCGGACTCTGCTGGCAGAGAACT 970
QY 897 GAACAAAGTAGTGAACAGCACCCGAGTACTGGGACACTGCCATCTCTTCAACAGGCA 956
DB 971 GAACAAAGTAGTGAACAGCACCCGAGTACTGGGACACTGCCATCTCTTCAACAGGCA 1030
QY 957 GGACCTGTGGGAGCCACCACTGTGACACCTGGGATGCTGATGGGTACCATGTG 1016
DB 1031 GGACCTGTGGGAGCCACCACTGTGACACCTGGGATGCTGATGGGTACCATGTG 1090
QY 1017 TGACCCCAAGAGAGCTCTCTGTCTATTGAGGAGATGGGCTTCATCAGCTTTCACAC 1076
DB 1091 TGACCCCAAGAGAGCTCTCTGTCTATTGAGGAGATGGGCTTCATCAGCTTTCACAC 1150
QY 1077 TGCCCAACAGCTGGGCGCAGCTGTTCACATGCCCATGCAATGTGAAGTCTGTGAGGA 1136
DB 1151 TGCCCAACAGCTGGGCGCAGCTGTTCACATGCCCATGCAATGTGAAGTCTGTGAGGA 1210
QY 1137 GGTGTTTGGGAGGCTCCCGAGCAACCACTGATGTCCCGGACCTCATCCAGATCGACCG 1196
DB 1211 GGTGTTTGGGAGGCTCCCGAGCAACCACTGATGTCCCGGACCTCATCCAGATCGACCG 1270
QY 1197 TGCCCAACCCCTGGTTCAGCTTCAGTGTGCTCATCATCAGGACTTCTTGGACAGTGGGCA 1256
DB 1271 TGCCCAACCCCTGGTTCAGCTTCAGTGTGCTCATCATCAGGACTTCTTGGACAGTGGGCA 1330
QY 1257 CGGTGACTGCTCTCTGGACCAACCCAGAGCCCATCTCCCTGCCCGAGTATCTGCGCGG 1316
DB 1331 CGGTGACTGCTCTCTGGACCAACCCAGAGCCCATCTCCCTGCCCGAGTATCTGCGCGG 1390
QY 1317 CGCCAGGTACACCTTGAGCCAGCAGTGGCTGGCTTTTGGGCTGGGCTCCAGAGCCCTG 1376
DB 1391 CGCCAGGTACACCTTGAGCCAGCAGTGGGCTGGCTTTTGGGCTGGGCTCCAGAGCCCTG 1450
QY 1377 TCTTTACATGAGTACTGACCAAGCTGTGGTGCACCCGGAGGCCAAGGAGCAGATGGT 1436
DB 1451 TCTTTACATGAGTACTGACCAAGCTGTGGTGCACCCGGAGGCCAAGGAGCAGATGGT 1510
QY 1437 GTGCCAGAGCCGCGACTTCCCTCGGGCGGATGGCAGCAGCTGTGGGAGGCGGAGCTCTG 1496
DB 1511 GTGCCAGAGCCGCGACTTCCCTCGGGCGGATGGCAGCAGCTGTGGGAGGCGGAGCTCTG 1570

QY	1497	CCTCAAGGGGCTGGTGGAGAGACACAACTCAACAGCACAGGTTGGATGTTCTCTG	1556	XX	Human metalloprotease partial DNA sequence #11.
Db	1571	CCTCAAGGGGCTGGTGGAGAGACACAACTCAACAGCACAGGTTGGATGTTCTCTG	1630	DE	
QY	1557	GGCCAAATGGATCCCTATGTGGCCCTGCTCGGCGACATGTGTGGGGCGTGACGTGGC	1616	KW	Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;
Db	1631	GGCCAAATGGATCCCTATGTGGCCCTGCTCGGCGACATGTGTGGGGCGTGACGTGGC	1690	KW	vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;
QY	1617	CAGAGGCACTGACCAACCCACCCCTGCCACCGGGGCAAGTACTGGAGGAGTGAG	1676	KW	hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;
Db	1691	CAGAGGCACTGACCAACCCACCCCTGCCACCGGGGCAAGTACTGGAGGAGTGAG	1750	KW	anorectic; antiinflammatory; aspartyl protease; cysteine protease;
QY	1677	GGTGAATACCGATCCTGCAACCTGAGCGCTGCCCGCCAGCTCCGGGAAGAGCTT	1736	KW	metalloprotease; serine protease; cancer; haematopoietic; breast; colon;
Db	1751	GGTGAATACCGATCCTGCAACCTGAGCGCTGCCCGCCAGCTCCGGGAAGAGCTT	1810	KW	lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;
QY	1737	CCGGAGGAGCAGTGTGAGGCTTTCAACGGCTACACACAGCACCAACCGGCTCACTCT	1796	KW	immune-related disease; cardiovascular disease; neuronal disease;
Db	1811	CCGGAGGAGCAGTGTGAGGCTTTCAACGGCTACACACAGCACCAACCGGCTCACTCT	1870	KW	migraine; sexual dysfunction; mood disorder; attention disorder;
QY	1797	CGCGTGGCATGGTCCCAAGTACTCCGCGGTGCTCCCGGGAAGTGCAGCTCAT	1856	KW	cognition disorder; hypotension; hypertension; psychotic disorder;
Db	1871	CGCGTGGCATGGTCCCAAGTACTCCGCGGTGCTCCCGGGAAGTGCAGCTCAT	1930	KW	dyskinesia; metabolic disorder; inflammatory disorder; ss.
QY	1857	CTGCCGAGCCATGGCACTGGCTACTTCTATGTCTGGCACCCAA---GGTGGTGACCG	1913	OS	Homo sapiens.
Db	1931	CTGCCGAGCCATGGCACTGGCTACTTCTATGTCTGGCACCCAA---GGTGGTGACCG	1990	XX	WO200183782-A2.
QY	1914	CACGCTGTCTCTGCTGACTCCAGCTCCGCTGTGTCCAGGCAAGTGCATCAAGCTGG	1973	XX	08-NOV-2001.
Db	1991	CACGCTGTCTCTGCTGACTCCAGCTCCGCTGTGTCCAGGCAAGTGCATCAAGCTGG	2050	XX	04-MAY-2001; 2001WO-US14431.
QY	1974	CTGTGATGGAACTCGGCTCCAAAGAGATTCGACAACTGTGGGTGTGGGGGAGA	2033	PF	04-MAY-2000; 2000US-201879P.
Db	2051	CTGTGATGGAACTCGGCTCCAAAGAGATTCGACAACTGTGGGTGTGGGGGAGA	2110	PR	(SUGEN-) SUGEN INC.
QY	2034	CAATAAGAGCTGAAGAAGTGACTGGACTCTTCAACCAAGCCCATGATGGCTACAAATTT	2093	XX	Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
Db	2111	CAATAAGAGCTGAAGAAGTGACTGGACTCTTCAACCAAGCCCATGATGGCTACAAATTT	2170	PI	Payne V;
QY	2094	CGTGGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCCGCCAGCGGTTACAAAGG	2153	XX	WPI; 2002-041502/05.
Db	2171	CGTGGTGGCCATCCCGCAGCGGCTCAAGCATCGACATCCGCCAGCGGTTACAAAGG	2230	DR	P-PSDB; AAU72899.
QY	2154	GCTGATCGGGATGACAACTACTGGCTCTGAAGACAGCAAGCAAGTACCTGCTCAA	2213	DR	Novel protease polypeptide useful for screening for substances that may
Db	2231	GCTGATCGGGATGACAACTACTGGCTCTGAAGACAGCAAGCAAGTACCTGCTCAA	2290	XX	be used to treat, e.g., cancers, immune-related diseases,
QY	2214	CGGGCATTTCTGTGTGGGGTGGAGCGGGACCTGGTGAAGGGCAGTCTGTGGCG	2273	CC	cardiovascular disease, migraine, pain, psychotic and inflammatory
Db	2291	CGGGCATTTCTGTGTGGGGTGGAGCGGGACCTGGTGAAGGGCAGTCTGTGGCG	2350	CC	disorders
QY	2274	GTACAGCGGCACGGCACAGCGGTGGAGCGCTTCCGGGCCATCTGGAGCC	2333	CC	Claim 30; Figure 1P; 232pp; English.
Db	2351	GTACAGCGGCACGGCACAGCGGTGGAGCGCTTCCGGGCCATCTGGAGCC	2410	XX	The invention relates to an isolated, enriched, or purified protease
QY	2334	GCTGACCTGGAGGTCTCTCCGTGGGGAAGATGACACCCCGCGGTCTGCTACTCCTT	2393	CC	polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
Db	2411	GCTGACCTGGAGGTCTCTCCGTGGGGAAGATGACACCCCGCGGTCTGCTACTCCTT	2470	CC	screen for substances (S) that may modulate its activity. Administering
QY	2394	CTATCTGCCCAAGAGCCTCGGGAGGACAAAGTCTCTCATCCC	2436	CC	S (which modulates protease activity in vitro) may be used to treat a
Db	2471	CTATCTGCCCAAGAGCCTCGGGAGGACAAAGTCTCTCATCCC	2513	CC	disease or disorder selected from cancers (e.g., of tissues, of blood or
RESULT 5				CC	haematopoietic origin, of the breast, colon, lung, prostate, cervical,
AAS97182				CC	brain, ovarian, bladder or kidney), immune-related diseases and
ID	AAS97182	standard; cDNA; 2937 BP.		CC	disorders, cardiovascular disease, brain or neuronal-associated diseases
XX				CC	(e.g., central or peripheral nervous system diseases, migraine, pain,
XX	AAS97182;			CC	sexual dysfunction, mood disorders, attention disorders, cognition
XX				CC	disorders, hypotension, hypertension, psychotic disorders, neurological
DT	26-FEB-2002 (first entry)			CC	disorders and dyskinesias), metabolic disorders and inflammatory

QY 237 CTCACCTGAGCATCTGGGCTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGGC 296
Db 312 CTCACCTGAGCATCTGGGCTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGGC 371
QY 297 ACCTGCTTCTATTCTGGGAGCGTGAACCCCGAGCCGGACTCTGCTGCTGTGAGCCT 356
Db 372 ACCTGCTTCTATTCTGGGAGCGTGAACCCCGAGCCGGACTCTGCTGCTGTGAGCCT 431
QY 357 GTCCGGGGGGCTCCCGGGAGCCTTTGGCTACCGAGCCCGGAGTAGTATGTCATTAGCCCGCT 416
Db 432 GTCCGGGGGGCTCCCGGGAGCCTTTGGCTACCGAGCCCGGAGTAGTATGTCATTAGCCCGCT 491
QY 417 GCCCAATGTAGCGCGCGCGCGCGAGCGCAACAGCCAGGCGCACACCTTCTCCAGCG 476
Db 492 GCCCAATGTAGCGCGCGCGCGCGAGCGCAACAGCCAGGCGCACACCTTCTCCAGCG 551
QY 477 CCGGGGTGTTCCGGGGGGGCTTCCGGAGACCCACCTCTCGCTCGGGGGTGGCCTCGGG 536
Db 552 CCGGGGTGTTCCGGGGGGGCTTCCGGAGACCCACCTCTCGCTCGGGGGTGGCCTCGGG 611
QY 537 CTGGAACCCCGCCATCTACGGGCGCTTGACCCCTTACAAGCGCGCGCGGGCTTCGG 596
Db 612 CTGGAACCCCGCCATCTACGGGCGCTTGACCCCTTACAAGCGCGCGGGGCTTCGG 671
QY 597 GGAGAGTCGTAGCGCGCGAGCTTGGCGGCGCAAGCGTTTCGTGCTATCCCGCGGTA 656
Db 672 GGAGAGTCGTAGCGCGCGAGCTTGGCGGCGCAAGCGTTTCGTGCTATCCCGCGGTA 731
QY 657 COTGGAGAGCTGTGTGTCGGGAGAGTCAATGTCAAGTTCACGGCGCGGAGCTTGG 716
Db 732 COTGGAGAGCTGTGTGTCGGGAGAGTCAATGTCAAGTTCACGGCGCGGAGCTTGG 791
QY 717 ACATTATCTGTGACGCTGCTGGCAACCGCGCGGAGCTTACCGCCATCCAGCATCCT 776
Db 792 ACATTATCTGTGACGCTGCTGGCAACCGCGCGGAGCTTACCGCCATCCAGCATCCT 851
QY 777 CAACCCCATCAACATCTGTGTGTCAGAGTCTGCTTCTTAGAGATCGTGAATCCCGGCG 836
Db 852 CAACCCCATCAACATCTGTGTGTCAGAGTCTGCTTCTTAGAGATCGTGAATCCCGGCG 911
QY 837 CAAGTCTACCGGCAATCGGCGCTGACGCTGCGCAACTTCTGTGCTGGCAAGAGAGCT 896
Db 912 CAAGTCTACCGGCAATCGGCGCTGACGCTGCGCAACTTCTGTGCTGGCAAGAGAGCT 971
QY 897 GAACAAAGTGAAGTGAACAGCACCAGGAGTACTGGGAGACTGCCATCTCTTCAACAGGCA 956
Db 972 GAACAAAGTGAAGTGAACAGCACCAGGAGTACTGGGAGACTGCCATCTCTTCAACAGGCA 1031
QY 957 GGACCTGTGTGAGGAGCACCCTGTGACACCCCTGGGATGCTCATGGGTACCATGTG 1016
Db 1032 GGACCTGTGTGAGGAGCACCCTGTGACACCCCTGGGATGCTCATGGGTACCATGTG 1091
QY 1017 TGACCCCAAGAGAGTCTGTCTGTCATTGAGGAGCTGGGCTTCCATCAGCCTTCAACAC 1076
Db 1092 TGACCCCAAGAGAGTCTGTCTGTCATTGAGGAGCTGGGCTTCCATCAGCCTTCAACAC 1151
QY 1077 TGCCCAAGAGTGGGAGCAGCTGTTCAACATGCCCCCATGACAAATGTGAAAGTCTGTGAGGA 1136
Db 1152 TGCCCAAGAGTGGGAGCAGCTGTTCAACATGCCCCCATGACAAATGTGAAAGTCTGTGAGGA 1211
QY 1137 GGTGTTTGGGAAGTCCGAGGAGCAACCATGATGTCCCGGAGCTCATCCAGATCGACCG 1196
Db 1212 GGTGTTTGGGAAGTCCGAGGAGCAACCATGATGTCCCGGAGCTCATCCAGATCGACCG 1271
QY 1197 TGCCCAACCCCTGTGAGCTGCTGTCATTGAGGAGCTGGGCTTCCATCAGCCTTCAACAC 1256
Db 1272 TGCCCAACCCCTGTGAGCTGCTGTCATTGAGGAGCTGGGCTTCCATCAGCCTTCAACAC 1331
QY 1257 CGGTGACTGCTCTCTGGACCAACCCAGCAGCCATCTCCCTCCCGAGAGTCTGCGGG 1316
Db 1332 CGGTGACTGCTCTCTGGACCAACCCAGCAGCCATCTCCCTCCCGAGAGTCTGCGGG 1391
QY 1317 GCGCAGCTACACCTGTAGCCAGCAGTGCAGCTGGCTTTTGGCGTGGCTTCCAGCCCTG 1376

Db 1392 CGCCAGCTACACCTGAGCCAGCAGTGCAGCTGGCTTTTGGCGTGGCTTCCAGCCCTG 1451
QY 1377 TCCCTTACATGACGACTGACCAAGCTGTGTGACCGGAGGAGCCCAAGGACAGATGGT 1436
Db 1452 TCCCTTACATGACGACTGACCAAGCTGTGTGACCGGAGGAGCCCAAGGACAGATGGT 1511
QY 1437 GTGCCAGACCCGCACTTCCCTTGGGCGATGGCACCAGCTGTGTGCGAGGAGGCAAGCTCTG 1496
Db 1512 GTGCCAGACCCGCACTTCCCTTGGGCGATGGCACCAGCTGTGTGCGAGGAGGCAAGCTCTG 1571
QY 1497 CCTCAAGGGGCGCTGCGTGGAGAGACACACCTCAACAGCAGGAGGAGTGTGCTTCTG 1556
Db 1572 CCTCAAGGGGCGCTGCGTGGAGAGACACACCTCAACAGCAGGAGGAGTGTGCTTCTG 1631
QY 1557 GGCCAATGGGATCCCTTATGGCCCTCTCGCGCACATGTGTGGGGGCTGCGAGCTGGC 1616
Db 1632 GGCCAATGGGATCCCTTATGGCCCTCTCGCGCACATGTGTGGGGGCTGCGAGCTGGC 1691
QY 1617 CAGGAGCAGTGCACCAACCCCTGCGCAAGGGGGCAAGTACTTGCAGGAGTGTGAG 1676
Db 1692 CAGGAGCAGTGCACCAACCCCTGCGCAAGGGGGCAAGTACTTGCAGGAGTGTGAG 1751
QY 1677 GGTGAAATACCGATCTGCAACCTGGAGCCCTGCGCCAGCTACAGCTTCCGGAAGAGCTT 1736
Db 1752 GGTGAAATACCGATCTGCAACCTGGAGCCCTGCGCCAGCTACAGCTTCCGGAAGAGCTT 1811
QY 1737 CCGGGAGGAGCAGTGTGAGGCTTTCAACGCTTCAACGCTTCAACGCTTCAACGCTTCACTCT 1796
Db 1812 CCGGGAGGAGCAGTGTGAGGCTTTCAACGCTTCAACGCTTCAACGCTTCAACGCTTCACTCT 1871
QY 1797 CCGCTGGCATGGTGGCCCAAGTACTCCGGGTGTCTCCCGGGAGCAAGTGCAGTCTCAT 1856
Db 1872 CCGCTGGCATGGTGGCCCAAGTACTCCGGGTGTCTCCCGGGAGCAAGTGCAGTCTCAT 1931
QY 1857 CTCGGAGGCAATGGCACTGGCTTCTTATGTGCTGGCACCCCAA --- GGTGGGAGCAGG 1913
Db 1932 CTCGGAGGCAATGGCACTGGCTTCTTATGTGCTGGCACCCCAAAGTGGT---GGACGG 1991
QY 1914 CAGCTGTGCTCTCTGACCTCCAGCTCCGCTGTGTGCTTCCAGGCAAGTGCATCAGGCTGG 1973
Db 1992 CAGCTGTGCTCTCTGACCTCCAGCTCCGCTGTGTGCTTCCAGGCAAGTGCATCAGGCTGG 2051
QY 1974 CTGTATGGGAACCTGGGCTTCCAGAGAGATTCGACAAAGTGTGGGTGTGTGGGGGAGA 2033
Db 2052 CTGTATGGGAACCTGGGCTTCCAGAGAGATTCGACAAAGTGTGGGTGTGTGGGGGAGA 2111
QY 2034 CAATAGAGCTGCAAGAGTGTGAGCTTCTTCAACAAAGCCCATGCTATGGCTTCAATTT 2093
Db 2112 CAATAGAGCTGCAAGAGTGTGAGCTTCTTCAACAAAGCCCATGCTATGGCTTCAATTT 2171
QY 2094 CGTGTGGCCATCCCGCAGGCGCTCAAGCATCGACATCCGCCAGCGGGTTACAAAG 2153
Db 2172 CGTGTGGCCATCCCGCAGGCGCTCAAGCATCGACATCCGCCAGCGGGTTACAAAG 2231
QY 2154 GCTGATCGGGATGACAACTACTTGGCTCTGAAGAACAGCCCAAGCAAGTACTGTCTAA 2213
Db 2232 GCTGATCGGGATGACAACTACTTGGCTCTGAAGAACAGCCCAAGCAAGTACTGTCTAA 2291
QY 2214 CGGGCATTTTGTGGTGTCTCGGCGGTGGAGGAGCCTGTGTGTGAAGGCAAGTGTCTGCTGG 2273
Db 2292 CGGGCATTTTGTGGTGTCTCGGCGGTGGAGGAGCCTGTGTGTGAAGGCAAGTGTCTGCTGG 2351
QY 2274 GTACAGCGCAGCGGCACAGCGGTGGAGAGCTTCAAGGCTTCCCGGCCCATCTCTGAGCG 2333
Db 2352 GTACAGCGCAGCGGCACAGCGGTGGAGAGCCTTCAAGGCTTCCCGGCCCATCTCTGAGCG 2411
QY 2334 GCTGACCGTGGAGTCTCTCTCGTGGGGAAGATGACACCGCCCGGGTCTCTACTCTCTT 2393
Db 2412 GCTGACCGTGGAGTCTCTCTCGTGGGGAAGATGACACCGCCCGGGTCTCTCTCTT 2471
QY 2394 CTATCTGCCCAAGAGCCTTCGGGAGGAGCAAGTCTCTCTATCCC 2436

```
Db 2472 CTATCTGCCAAAGAGCTCGGGAGGACAAAGTCTCTCATCCC 2514
RESULT 6
AAD35570
ID AAD35570 standard; cDNA; 1104 BP.
XX
AC AAD35570;
XX
XX 26-JUL-2002 (first entry)
XX
XX Human protease cDNA #3.
DE
XX Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme; gene; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX 1..1104
XX /tag= a
XX /product= "Human protease #3"
XX
XX WO200226949-A2.
XX
XX 04-APR-2002..
XX
XX 27-SEP-2001; 2001WO-US30350.
XX
XX 29-SEP-2000; 2000US-236689P.
XX
XX (LEXI-) LEXICON GENETICS INC.
XX
XX Friddle CJ, Hilbun E;
XX
XX WPI; 2002-372123/40.
XX
XX P-PSDB; AAE22542.
XX
XX Novel nucleic acid encoding a human protease, useful as a hybridization
XX probe for screening libraries and assessing gene expression patterns -
XX
XX Disclosure; Page 38-39; 41pp; English.
XX
XX The present sequence is a cDNA encoding novel human protein (NHP),
XX human protease. NHPs share structural similarity with animal proteases
XX particularly zinc metalloproteases. Sequences of the invention are
XX useful in therapeutic, diagnostic and pharmacogenomic applications.
XX NHP polynucleotides are used as hybridisation probes for screening
XX libraries and assessing gene expression patterns. They can also be
XX used for treating related biological disorders such as obesity, high
XX blood pressure, arthritis, connective tissue disorders and infertility.
XX They are also used in gene therapy..
XX
XX Sequence 1104 BP; 189 A; 357 C; 345 G; 213 T; 0 other;
XX
XX Query Match 36.3%; Score 1035; DB 24; Length 1104;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1035; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 57 TGAGCCAGAGCGGGAGTGTCTCCATCCGACTGGACCCGACATTAAACGGCCGCG 116
Db 57 TGAGCCAGAGCGGGAGTGTCTCCATCCGACTGGACCCGACATTAAACGGCCGCG 116
QY 117 CTACTACTGGGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCACAGC 176
Db 117 CTACTACTGGGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCACAGC 176
QY 177 ATTTTCAGAGGACTTTTACCTACACCTGACCCGGATCTCAGTCTTGGCTCCGCGCTT 236
Db 177 ATTTTCAGAGGACTTTTACCTACACCTGACCCGGATCTCAGTCTTGGCTCCGCGCTT 236
QY 237 CTCCACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTACCCGGGGCTCTTCAGACCTGCG 296
Db 237 CTCCACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTCTTCAGACCTGCGCG 296
QY 297 ACCTGTCTTCTATTCTGGGAGCTGAACCGCGAGCCGACTCGTTCGCTGTGAGCCT 356
Db 297 ACCTGTCTTCTATTCTGGGAGCTGAACCGCGAGCCGACTCGTTCGCTGTGAGCCT 356
QY 357 GTGCGGGGGTCCGCGAGCCTTTGGCTACCGAGGCGCGAGTATGTATATACCCGCT 416
Db 357 GTGCGGGGGTCCGCGAGCCTTTGGCTACCGAGGCGCGAGTATGTATATAGCGCT 416
QY 417 GCCCAATCTAGCGCGCGCGGCGAGCGCAACAGCCAGGCGCACACCTCTTCAGCG 476
Db 417 GCCCAATCTAGCGCGCGCGGCGAGCGCAACAGCCAGGCGCACACCTCTTCAGCG 476
QY 477 CGGGGGTGTTCGCGGCGCGCTTCCGAGAGCCCAACCTCTCGCTCGGGGTGCTCGGG 536
Db 477 CGGGGGTGTTCGCGGCGCGCTTCCGAGAGCCCAACCTCTCGCTCGGGGTGCTCGGG 536
QY 537 CTGGAACCCCGCATCTACGGGCGCTTACAGCCGCGGCGCGGGGGGCTTCGG 596
Db 537 CTGGAACCCCGCATCTACGGGCGCTTACAGCCGCGGCGCGGGGGGCTTCGG 596
QY 597 GGAGAGTCTAGCGCGCGAGTCTGGGCGCGGCGGCGGCGGCGGCGGTA 656
Db 597 GGAGAGTCTAGCGCGCGAGTCTGGGCGCGGCGGCGGCGGCGGCGGTA 656
QY 657 CGTGAGAGCGTGTGGTGGGAGCGAGTCAATGGTCAAGTTCACGCGCGGAGCTGGA 716
Db 657 CGTGAGAGCGTGTGGTGGGAGCGAGTCAATGGTCAAGTTCACGCGCGGAGCTGGA 716
QY 717 ACATTATCTGTGACGCTGTGGCAACGCGCGGCGGCGGCGGCGGCGGATCCT 776
Db 717 ACATTATCTGTGACGCTGTGGCAACGCGCGGCGGCGGCGGCGGCGGATCCT 776
QY 777 CAACCCCATCAACATCGTGTGCTCAAGTGTCTTCTTAGAGATCGTACTCCGGGCG 836
Db 777 CAACCCCATCAACATCGTGTGCTCAAGTGTCTTCTTAGAGATCGTACTCCGGGCG 836
QY 837 CAAGGTACCGGCAATGCGGCGCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAAGCT 896
Db 837 CAAGGTACCGGCAATGCGGCGCTGACGCTGCGCAACTTCTGTGCTGGCAGAGAAGCT 896
QY 897 GAACAAAGTGTGACAAAGCAACCCGAGTGTGGGAGCTGCTATCTTCACCAAGCA 956
Db 897 GAACAAAGTGTGACAAAGCAACCCGAGTGTGGGAGCTGCTATCTTCACCAAGCA 956
QY 957 GGACCTGTGTGGAGCAACCACTGTGACACCTCGGCGATGGCTGATGGGTACCATGTG 1016
Db 957 GGACCTGTGTGGAGCAACCACTGTGACACCTCGGCGATGGCTGATGGGTACCATGTG 1016
QY 1017 TGACCCCAAGAGAGCTGTCTGTCATTGAGGAGGCTTCCATCAGCTTCACCAAC 1076
Db 1017 TGACCCCAAGAGAGCTGTCTGTCATTGAGGAGGCTTCCATCAGCTTCACCAAC 1076
QY 1077 TGCCCAAGAGCTGGG 1091
Db 1077 TGCCCAAGAGCTGGG 1091
RESULT 7
AAD35568
ID AAD35568 standard; cDNA; 966 BP.
XX
AC AAD35568;
XX
XX 26-JUL-2002 (first entry)
XX
XX Human protease cDNA #1..
DE
XX Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme; gene; ss.
```


DE	Rat metalloproteinase ADAMTS-5 cDNA.
XX	
KW	Rat: ADAMTS-5; metalloproteinase; ADAM;
KW	a disintegrin and metalloproteinase domain; thrombospondin domain;
KW	vaccine; nontropic; neuroprotective; anti-parkinsonian;
KW	ceroprotective; cytostatic; antiarthritic; immunosuppressive;
KW	Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW	autoimmune disease; brain tumour; brain injury; ss.
XX	
OS	Rattus norvegicus.
XX	
FF	Key Location/Qualifiers
FT	CDS 1..1515
FT	/*tag= a
FT	/partial
FT	/product= "ADAMTS-5"
XX	
PN	WO200053774-A2.
XX	
PD	14-SEP-2000.
XX	
XX	08-MAR-2000; 2000WO-US06237.
PF	
XX	
XX	08-MAR-1999; 99US-0264585.
PR	
XX	
PA	(NEUR-) NEUROCRINE BIOSCIENCES INC.
XX	
PI	Kelner GS, Clark M, Maki RA;
PI	
DR	WPI: 2000-594326/56.
DR	P-PSDB; A821257.
XX	
PT	Polynucleotide encoding novel members of a disintegrin,
PT	metalloproteinase and thrombospondin domain protein family used to
PT	prevent and treat Alzheimer's disease, cancer and autoimmune diseases
XX	
XX	Claim 2; Fig 13; 129pp; English.
PS	
XX	
CC	The present sequence encodes rat metalloproteinase ADAMTS-5. The
CC	ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
CC	and Metalloproteinase Domain) family. Members of the ADAMTS family
CC	contain a thrombospondin domain in addition to the disintegrin and
CC	metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
CC	useful for the manufacture of medicaments for treating conditions
CC	associated with neuroinflammation and/or neurodegeneration, such as
CC	Alzheimer's disease, Parkinson's disease and stroke. They are also
CC	useful for treating conditions associated with cell proliferation, cell
CC	migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC	and autoimmune diseases. They can be used to treat patients afflicted
CC	with an invasive tumour, a brain tumour or brain injury.
XX	
SQ	Sequence 1518 BP; 307 A; 470 C; 462 G; 279 T; 0 other;
	Query Match 1.8%; Score 50; DB 21; Length 1518;
	Best Local Similarity 100.0%; Pred. No. 8.6e-13;
	Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps
Qy	1180 CTATCCAGATCGACGTCGCCAACCCCTGGTCAGCGTCAGTGTGCCAT 1229
Db	952 CTATCCAGATCGACGTCGCCAACCCCTGGTCAGCGTCAGTGTGCCAT 1001
RESULT 10	
AAH41029/c	
ID	AAH41029 standard; DNA; 30 BP.
XX	
AC	AAH41029;
XX	
XX	
DT	23-AUG-2001 (first entry)
XX	
DE	Adaptor primer SEQ ID 22 used in metalloproteinase DNA isolation.
XX	
KW	Metalloproteinase; human; aggrecanase; joint disease; osteoarthritis;

```

KW osteopathic; antiarthritic; primer; adaptor; ss.
OS Homo sapiens.
XX WO200134785-A1.
PN 17-MAY-2001.
XX 10-NOV-2000; 2000WO-JP07917.
XX 11-NOV-1999; 99JP-0321740.
PR 16-MAY-2000; 2000JP-0144020.
XX (YAMA ) YAMANOUCHI PHARM CO LTD.
PA (KAZU-) KAZUSA DNA RES INST.
XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
XX WPI; 2001-343602/36.
XX Metalloprotease with aggrecanase activity for treating joint diseases
PT especially osteoarthritis -
XX Example 13; Page 66; 85pp; Japanese.
CC This invention relates to a metalloprotease with aggrecanase activity.
CC The invention includes protein and DNA sequences of the metalloprotease,
CC vectors containing the DNA, host cells transformed by the vectors, and
CC antibodies directed against the metalloprotease. The antibodies, protein
CC and DNA sequences can be used in the treatment and prevention of joint
CC diseases, particularly osteoarthritis. The treatment may result in
CC osteopathic and antiarthritic activity. The present sequence represents
CC an adaptor primer used in the isolation and characterisation of the
CC metalloprotease gene of the invention.
XX SQ Sequence 30 BP; 8 A; 6 C; 10 G; 6 T; 0 other;

Query Match 1.1%; Score 30; DB 22; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 TTTACCTACACCTGACCGCGGATGCTCAGT 220
DB 30 TTTACCTACACCTGACCGCGGATGCTCAGT 1

RESULT 11
AAH41030/C
ID AAH41030 standard; DNA; 30 BP.
AC AAH41030;
XX 23-AUG-2001 (first entry)
DE Adaptor primer SEQ ID 23 used in metalloprotease DNA isolation.
XX Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;
KW osteopathic; antiarthritic; primer; adaptor; ss.
XX Homo sapiens.
XX WO200134785-A1.
XX 17-MAY-2001.
XX 10-NOV-2000; 2000WO-JP07917.
XX 11-NOV-1999; 99JP-0321740.
PR 16-MAY-2000; 2000JP-0144020.
XX (YAMA ) YAMANOUCHI PHARM CO LTD.
PA (KAZU-) KAZUSA DNA RES INST.
XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
XX WPI; 2001-343602/36.
XX Metalloprotease with aggrecanase activity for treating joint diseases
PT especially osteoarthritis -
XX Example 13; Page 66; 85pp; Japanese.
CC This invention relates to a metalloprotease with aggrecanase activity.
CC The invention includes protein and DNA sequences of the metalloprotease,
CC vectors containing the DNA, host cells transformed by the vectors, and
CC antibodies directed against the metalloprotease. The antibodies, protein
CC and DNA sequences can be used in the treatment and prevention of joint
CC diseases, particularly osteoarthritis. The treatment may result in
CC osteopathic and antiarthritic activity. The present sequence represents
CC an adaptor primer used in the isolation and characterisation of the
CC metalloprotease gene of the invention.
XX SQ Sequence 30 BP; 8 A; 6 C; 10 G; 6 T; 0 other;

```

```

PI Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
XX WPI; 2001-343602/36.
XX Metalloprotease with aggrecanase activity for treating joint diseases
PT especially osteoarthritis -
XX Example 13; Page 66; 85pp; Japanese.
XX This invention relates to a metalloprotease with aggrecanase activity.
XX The invention includes protein and DNA sequences of the metalloprotease,
XX vectors containing the DNA, host cells transformed by the vectors, and
XX antibodies directed against the metalloprotease. The antibodies, protein
XX and DNA sequences can be used in the treatment and prevention of joint
XX diseases, particularly osteoarthritis. The treatment may result in
XX osteopathic and antiarthritic activity. The present sequence represents
XX an adaptor primer used in the isolation and characterisation of the
XX metalloprotease gene of the invention.
XX SQ Sequence 30 BP; 10 A; 6 C; 6 G; 8 T; 0 other;

Query Match 1.1%; Score 30; DB 22; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 161 TTTTTCAGATCACAGCATTTTCAGGAGGACT 190
DB 30 TTTTTCAGATCACAGCATTTTCAGGAGGACT 1

RESULT 12
AAH41017/C
ID AAH41017 standard; DNA; 41 BP.
XX AC AAH41017;
XX 23-AUG-2001 (first entry)
DE PCR primer specific for human metalloprotease DNA.
XX KW Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;
XX KW osteopathic; antiarthritic; PCR primer; ss.
XX OS Homo sapiens.
XX PN WO200134785-A1.
XX PD 17-MAY-2001.
XX 10-NOV-2000; 2000WO-JP07917.
XX 11-NOV-1999; 99JP-0321740.
PR 16-MAY-2000; 2000JP-0144020.
XX (YAMA ) YAMANOUCHI PHARM CO LTD.
PA (KAZU-) KAZUSA DNA RES INST.
XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
XX WPI; 2001-343602/36.
XX Metalloprotease with aggrecanase activity for treating joint diseases
PT especially osteoarthritis -
XX Example 4; Page 62; 85pp; Japanese.
XX This invention relates to a metalloprotease with aggrecanase activity.
XX The invention includes protein and DNA sequences of the metalloprotease,
XX vectors containing the DNA, host cells transformed by the vectors, and
XX antibodies directed against the metalloprotease. The antibodies, protein
XX and DNA sequences can be used in the treatment and prevention of joint
XX diseases, particularly osteoarthritis. The treatment may result in
XX osteopathic and antiarthritic activity. The present sequence represents
XX

```

CC a PCR primer used in the isolation and characterisation of the
XX metalloprotease gene of the invention.

SQ Sequence 41 BP; 5 A; 16 C; 13 G; 7 T; 0 other;

Query Match 1.1%; Score 30; DB 22; Length 41;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1720 GCCTCCGGAAGAGCTTCGGGAGGAGCAG 1749

Db 41 GCCTCCGGAAGAGCTTCGGGAGGAGCAG 12

RESULT 13

AAH41016
ID AAH41016 standard; DNA; 42 BP.

XX AC AAH41016;

DT 23-AUG-2001 (first entry)

DE PCR primer specific for human metalloprotease DNA.

KW Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;

KW osteopathic; antiarthritic; PCR primer; ss.

XX OS Homo sapiens.

XX PN WO200134785-A1.

XX PD 17-MAY-2001.

XX PF 10-NOV-2000; 2000WO-JP07917.

XX PR 11-NOV-1999; 99JP-0321740.

XX PR 16-MAY-2000; 2000JP-0144020.

XX PA (YAMA) YAMANOUCHI PHARM CO LTD.

XX PA (KAZU-) KAZUSA DNA RES INST.

XX PI Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;

XX WPI; 2001-343602/36.

XX Metalloprotease with aggrecanase activity for treating joint diseases
PT especially osteoarthritis -

PS Example 4; Page 62; 85pp; Japanese.

XX CC This invention relates to a metalloprotease with aggrecanase activity.
CC The invention includes protein and DNA sequences of the metalloprotease.
CC vectors containing the DNA, host cells transformed by the vectors, and
CC antibodies directed against the metalloprotease. The antibodies, protein
CC and DNA sequences can be used in the treatment and prevention of joint
CC diseases, particularly osteoarthritis. The treatment may result in
CC osteopathic and antiarthritic activity. The present sequence represents
CC a PCR primer used in the isolation and characterisation of the
CC metalloprotease gene of the invention.

XX SQ Sequence 42 BP; 6 A; 12 C; 11 G; 13 T; 0 other;

Query Match 1.1%; Score 30; DB 22; Length 42;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTTTTCTGGGCATCTACCTGGCT 30

Db 13 ATGCTTTTCTGGGCATCTACCTGGCT 42

RESULT 14

AAH41018

ID AAH41018 standard; DNA; 27 BP.

XX AC AAH41018;

DT 23-AUG-2001 (first entry)

XX PCR primer specific for human metalloprotease DNA SEQ ID 9.

DE Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;

KW osteopathic; antiarthritic; PCR primer; ss.

XX OS Homo sapiens.

XX PN WO200134785-A1.

XX PD 17-MAY-2001.

XX PF 10-NOV-2000; 2000WO-JP07917.

XX PR 11-NOV-1999; 99JP-0321740.

XX PR 16-MAY-2000; 2000JP-0144020.

XX PA (YAMA) YAMANOUCHI PHARM CO LTD.

XX PA (KAZU-) KAZUSA DNA RES INST.

XX PI Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;

XX WPI; 2001-343602/36.

XX Metalloprotease with aggrecanase activity for treating joint diseases
PT especially osteoarthritis -

PS Example 5; Page 63; 85pp; Japanese.

XX CC This invention relates to a metalloprotease with aggrecanase activity.
CC The invention includes protein and DNA sequences of the metalloprotease.
CC vectors containing the DNA, host cells transformed by the vectors, and
CC antibodies directed against the metalloprotease. The antibodies, protein
CC and DNA sequences can be used in the treatment and prevention of joint
CC diseases, particularly osteoarthritis. The treatment may result in
CC osteopathic and antiarthritic activity. The present sequence represents
CC a PCR primer used in the isolation and characterisation of the
CC metalloprotease gene of the invention.

XX SQ Sequence 27 BP; 5 A; 6 C; 11 G; 5 T; 0 other;

Query Match 0.9%; Score 27; DB 22; Length 27;

Best Local Similarity 100.0%; Pred. No. 0.042;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1534 AAGCACAGGGTGGATGCTCTGGGCC 1560

Db 1 AAGCACAGGGTGGATGCTCTGGGCC 27

RESULT 15

AAH41019/c

ID AAH41019 standard; DNA; 37 BP.

XX AC AAH41019;

DT 23-AUG-2001 (first entry)

XX PCR primer specific for human metalloprotease DNA SEQ ID 10.

DE Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;

KW osteopathic; antiarthritic; PCR primer; ss.

XX OS Homo sapiens.

XX PN WO200134785-A1.

XX PD 17-MAY-2001.


```

XX 10-NOV-2000; 2000WO-JP07917.
XX
XX 11-NOV-1999; 99JP-0321740.
XX
XX 16-MAY-2000; 2000JP-0144020.
XX
XX (YAMA ) YAMANOUCHI PHARM CO LTD.
XX (KAZU-) KAZUSA DNA RES INST.
XX
XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
XX
XX WPI: 2001-343602/36.
XX
XX Metalloprotease with aggrecanase activity for treating joint diseases
XX especially osteoarthritis
XX
XX Example 5; Page 63; 85pp; Japanese.
XX
XX This invention relates to a metalloprotease with aggrecanase activity.
XX The invention includes protein and DNA sequences of the metalloprotease,
XX vectors containing the DNA, host cells transformed by the vectors, and
XX antibodies directed against the metalloprotease. The antibodies, protein
XX and DNA sequences can be used in the treatment and prevention of joint
XX diseases, particularly osteoarthritis. The treatment may result in
XX osteopathic and antiarthritic activity. The present sequence represents
XX a PCR primer used in the isolation and characterisation of the
XX metalloprotease gene of the invention.
XX
XX Sequence 37 BP; 7 A; 14 C; 14 G; 2 T; 0 other;
SQ
Query Match 0.9%; Score 27; DB 22; Length 37;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2824 CTGGACTTCTGCGTCTGAGGCGGTGC 2850
Db 37 CTGGACTTCTGCGTCTGAGGCGGTGC 11

```

Search completed: May 2, 2003, 01:10:45
Job time : 622 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 1, 2003, 22:52:52 ; Search time 103 Seconds
(without alignments)
8494.649 Million cell updates/sec

Title: US-10-009-332-2
Perfect score: 2853
Sequence: 1 atgcttttgcggcctcct.....gcgtcctgagccgtgctga 2853

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	0.8	23	2114	4	US-09-130-491-7
2	0.7	19	4533	3	US-08-726-214-5
3	0.6	18	311	4	US-09-643-597-314
4	0.6	18	2457	4	US-08-872-757-1
5	0.6	18	2487	1	US-08-377-292-1
6	0.6	18	2830	2	US-09-010-928B-1
7	0.6	18	3226	1	US-07-862-021B-11
8	0.6	18	3226	1	US-08-313-288B-11
9	0.6	18	3226	5	PCT-US93-03164-11
10	0.6	18	3546	4	US-08-872-757-3
11	0.6	18	4149	2	US-08-737-715-1
12	0.6	18	4403765	4	US-09-103-840A-2
13	0.6	18	4403765	4	US-09-103-840A-2
14	0.6	18	4411529	4	US-09-103-840A-1
15	0.6	18	4411529	4	US-09-103-840A-1
16	0.6	17	426	4	US-08-914-375C-55
17	0.6	17	438	1	US-08-540-242A-5
18	0.6	17	438	5	PCT-US96-01471-5
19	0.6	17	444	1	US-08-688-609-7
20	0.6	17	444	3	US-09-002-832-7
21	0.6	17	445	4	US-08-688-908-6
22	0.6	17	504	1	US-08-540-242A-1
23	0.6	17	504	4	US-08-692-922-6
24	0.6	17	504	5	PCT-US96-01471-1
25	0.6	17	701	3	US-08-292-345B-1
26	0.6	17	702	2	US-09-003-081-1
27	0.6	17	702	3	US-08-648-262-1

28	17	0.6	702	3	US-08-648-263-1	Sequence 1, Appli
29	17	0.6	900	2	US-08-518-862C-5	Sequence 5, Appli
30	17	0.6	1205	3	US-09-120-772-1	Sequence 1, Appli
31	17	0.6	1452	1	US-08-187-785-2	Sequence 2, Appli
32	17	0.6	1480	4	US-09-142-569-1	Sequence 1, Appli
33	17	0.6	1659	1	US-08-333-358-7	Sequence 7, Appli
34	17	0.6	1659	1	US-08-463-694-7	Sequence 7, Appli
35	17	0.6	1659	1	US-08-694-501-7	Sequence 7, Appli
36	17	0.6	2023	4	US-09-288-143-54	Sequence 54, Appli
37	17	0.6	2793	2	US-08-347-563A-1	Sequence 1, Appli
38	17	0.6	2793	3	US-08-485-942A-1	Sequence 1, Appli
39	17	0.6	2793	3	US-08-488-214A-1	Sequence 1, Appli
40	17	0.6	2793	3	US-08-488-208A-1	Sequence 1, Appli
41	17	0.6	2793	4	US-08-483-211A-1	Sequence 1, Appli
42	17	0.6	2793	4	US-08-488-223A-1	Sequence 1, Appli
43	17	0.6	2793	4	US-08-438-431A-1	Sequence 1, Appli
44	17	0.6	2855	2	US-08-852-153-1	Sequence 1, Appli
45	17	0.6	3255	2	US-08-852-153-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-130-491-7
; Sequence 7, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-75, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 2114
; TYPE: DNA
; ORGANISM: Rattus rattus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(1445)
US-09-130-491-7

Query Match 0.8%; Score 23; DB 4; Length 2114;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2008 GACAACTGTGGGTGTGTGGGG 2030
Db 807 GACAACTGTGGGTGTGTGGGG 829
|||||

RESULT 2
US-08-726-214-5
; Sequence 5, Application US/08726214
; Patent No. 6107076
; GENERAL INFORMATION:
; APPLICANT: Tang, Wei-Jen
; APPLICANT: Gilman, Alfred G.
; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
; TITLE OF INVENTION: AND USES THEREFOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas

COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,214
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/005,498
FILING DATE: 04-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: UTSD:450
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4533 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-726-214-5

Query Match 0.7%; Score 19; DB 3; Length 4533;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1986 CCTGGGCTCCAAGAAGAGA 2004
|||||
DB 3078 CCTGGGCTCCAAGAAGAGA 3096

RESULT 3
US-09-643-597-314/c
; Sequence 314, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 314
; LENGTH: 311
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-643-597-314

Query Match 0.6%; Score 18; DB 4; Length 311;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1988 TGGGCTCCAAGAAGAGAT 2005
|||||

Db 86 TGGGCTCCAAGAAGAGAT 69
RESULT 4
US-08-872-757-1
; Sequence 1, Application US/08872757
; Patent No. 6258584
; GENERAL INFORMATION:
; APPLICANT: Prockop, Darwin J.
; APPLICANT: Hojima, Yoshio
; APPLICANT: Li, Shi-Wu
; APPLICANT: Sieron, Aleksander
; TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND
; TITLE OF INVENTION: PROCESSES; METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/872,757
; FILING DATE: 10-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,187
; FILING DATE: 01-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Halluin, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: 8389-028-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-854-3660
; TELEFAX: 415-854-3694
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2457 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2190
US-08-872-757-1
Query Match 0.6%; Score 18; DB 4; Length 2457;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1080 CCACGAGCTGGGCCACGT 1097
|||||
DB 636 CCACGAGCTGGGCCACGT 653
RESULT 5
US-08-377-292-1
; Sequence 1, Application US/08377292
; Patent No. 5693615
; GENERAL INFORMATION:
; APPLICANT: STONE, ROGER L.
; TITLE OF INVENTION: THERAPEUTIC FORMULAS FOR OSTEOINDUCTION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company

```

; STREET: 11810 East Miami River Road
; CITY: Cincinnati
; STATE: Ohio
; COUNTRY: U.S.A.
; ZIP: 45239-8707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/377,292
; FILING DATE: 23-JAN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/243,435
; FILING DATE:
; APPLICATION NUMBER: US/08/117,367
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Corstanje, Brahm J.
; REGISTRATION NUMBER: 34,804
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 513-245-2858
; TELEFAX: 513-741-3012
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2487 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-377-292-1

```

```

Query Match 0.6%; Score 18; DB 1; Length 2487;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1080 CCACGAGCTGGCCACGT 1097
Db 665 CCACGAGCTGGCCACGT 682

```

```

RESULT 6
US-09-010-928B-1
; Sequence 1, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V
; APPLICANT: Hayashi, Cheryl Y
; TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
; TITLE OF INVENTION: CODING THEREFOR
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 8110 GATEHOUSE RD. SUITE 500E
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,928B
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28977
; REFERENCE/DOCKET NUMBER: 1447-109P

```

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2830 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..2830
; OTHER INFORMATION: /note= "Flagelliform DNA sequence
; OTHER INFORMATION: taken from the 5' region. The putative start codon is at
; OTHER INFORMATION: position 219"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 219..2830
; US-09-010-928B-1
; Query Match 0.6%; Score 18; DB 2; Length 2830;
; Best Local Similarity 100.0%; Pred. No. 75;
; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; QY 2234 CGGTGAGCGGGACCTGG 2251
; Db 2384 CGGTGAGCGGGACCTGG 2401
; RESULT 7
; US-07-862-021B-11
; Sequence 11, Application US/07862021B
; Patent No. 5279966
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Kiar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,021B
; FILING DATE: 19920405
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3226 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 136..2543
; US-07-862-021B-11

```

```
Query Match          0.6%; Score 18; DB 1; Length 3226;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1203 CCCCTGGTCAGCCTGCAG 1220
      |||||
Db 1467 CCCCTGGTCAGCCTGCAG 1484

RESULT 8
PCT-US93-313-288B-11
; Sequence 11, Application US/08313288B
; Patent No. 5750502
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M. and Avihu Klar
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/313,288B
; FILING DATE: January 5, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0526
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3226 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 136..2543
; US-08-313-288B-11

Query Match          0.6%; Score 18; DB 1; Length 3226;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1203 CCCCTGGTCAGCCTGCAG 1220
      |||||
Db 1467 CCCCTGGTCAGCCTGCAG 1484

RESULT 9
PCT-US93-03164-11
; Sequence 11, Application PC/TUS9303164
; GENERAL INFORMATION:
; APPLICANT: Jessell, Thomas M
; APPLICANT: Klar, Avihu
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES OF A
; TITLE OF INVENTION: NOVEL SECRETED PROTEIN, F-SPONDIN
; NUMBER OF SEQUENCES: 20
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03164
; FILING DATE: 19930402
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3226 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 136..2543
; PCT-US93-03164-11

Query Match          0.6%; Score 18; DB 5; Length 3226;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1203 CCCCTGGTCAGCCTGCAG 1220
      |||||
Db 1467 CCCCTGGTCAGCCTGCAG 1484

RESULT 10
US-08-872-757-3
; Sequence 3, Application US/08872757
; Patent No. 6258584
; GENERAL INFORMATION:
; APPLICANT: Prockop, Darwin J.
; APPLICANT: Hojima, Yoshio
; APPLICANT: Li, Shi-Wu
; APPLICANT: Sieron, Aleksander
; TITLE OF INVENTION: RECOMBINANT C-PROTEINASE AND
; TITLE OF INVENTION: PROCESSES; METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/872,757
FILING DATE: 10-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/609,187
FILING DATE: 01-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 8389-028-999
TELEPHONE: 415-854-3660
TELEFAX: 415-854-3694
TELEX: 66141 PENNTE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3546 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2958
US-08-872-757-3

Query Match 0.6%; Score 18; DB 4; Length 3546;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1080 CCACGAGCTGGCCAGCT 1097
|||||

Db 636 CCACGAGCTGGCCAGCT 653

RESULT 11

US-08-737-715-1/c
Sequence 1, Application US/08/737,715
Patent No. 5958685
GENERAL INFORMATION:
APPLICANT: Ebina, Yousuke
TITLE OF INVENTION: MUTANT HUMAN INSULIN RECEPTOR DNA
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: SUGHRUE, MION, ZINN, MACPEAK & SEAS
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,715
FILING DATE: 12-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 134827/1995
FILING DATE: 12-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Nakamura, Dean H.
REGISTRATION NUMBER: 33,981
REFERENCE/DOCKET NUMBER: Q-43323
TELEPHONE: (202)293-7060
TELEFAX: (202)293-7860
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4149 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4146
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 82..4146
US-08-737-715-1

Query Match 0.6%; Score 18; DB 2; Length 4149;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1701 GGAGCCCTGCCCGAGCTC 1718
|||||

Db 3099 GGAGCCCTGCCCGAGCTC 3082

RESULT 12

US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 0.6%; Score 18; DB 4; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 433 CGGCGGCGGCGGCGCAAC 450
|||||

Db 77489 CGGCGGCGGCGGCGCAAC 77506

RESULT 13

US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765

; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 0.6%; Score 18; DB 4; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2225 TGGTGTGCGGGTGGAGC 2242
|||||
DB 3375303 TGGTGTGCGGGTGGAGC 3375286

RESULT 14
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 0.6%; Score 18; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 433 CCGGCGGCGGCGCAAC 450
|||||
DB 77510 CCGGCGGCGGCGCAAC 77527

RESULT 15
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 0.6%; Score 18; DB 4; Length 4411529;

Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2225 TGGTGTGCGGGTGGAGC 2242
|||||
DB 3379611 TGGTGTGCGGGTGGAGC 3379594

Search completed: May 2, 2003, 05:13:52
Job time : 14589 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic : nucleic search, using sw model

Run on: May 2, 2003, 00:35:12 ; Search time 339 Seconds
(without alignments)
9944.440 Million cell updates/sec

Title: US-10-009-332-2
Perfect score: 2853
Sequence: 1 atgtctttgtggcatcct.....gcgtcctgagccgtgcta 2853

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 746064 seqs, 590810554 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCNT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCNTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2695	94.5	2853	10 US-09-965-631-3	Sequence 3, Appli
2	2695	94.5	3446	10 US-09-965-631-7	Sequence 7, Appli
3	1435	50.3	2469	9 US-10-163-316-3	Sequence 3, Appli
4	1435	50.3	2940	9 US-10-163-316-1	Sequence 1, Appli
5	1035	36.3	1104	10 US-09-965-631-5	Sequence 5, Appli
6	903	31.7	966	10 US-09-965-631-1	Sequence 1, Appli
7	23	0.8	2114	12 US-10-105-929-7	Sequence 1, Appli
8	22	0.8	1813	9 US-09-931-457A-10	Sequence 10, Appl
9	21	0.7	506	9 US-09-918-995-22019	Sequence 22019, A
10	21	0.7	1827	10 US-09-925-297-180	Sequence 180, Appl
11	20	0.7	1017	9 US-09-938-842A-60	Sequence 60, Appl
12	19	0.7	256	9 US-09-535-459-1835	Sequence 1835, Ap
13	19	0.7	408	9 US-09-918-995-15962	Sequence 15962, A
14	19	0.7	431	10 US-09-925-297-85	Sequence 85, Appl
15	19	0.7	452	9 US-09-918-995-32286	Sequence 32286, A
16	19	0.7	466	9 US-09-918-995-34926	Sequence 34926, A
17	19	0.7	469	9 US-09-918-995-11410	Sequence 11410, A
18	18	0.6	157	10 US-09-864-761-19861	Sequence 19861, A
19	18	0.6	258	10 US-09-864-761-29711	Sequence 29711, A

c 20	18	0.6	311	10 US-09-735-705-314	Sequence 314, App
c 21	18	0.6	311	10 US-09-850-716A-314	Sequence 314, App
c 22	18	0.6	311	10 US-09-897-778-314	Sequence 314, App
c 23	18	0.6	381	10 US-09-878-574-669	Sequence 669, App
c 24	18	0.6	401	9 US-09-954-531-91	Sequence 91, Appli
c 25	18	0.6	401	10 US-09-954-456-752	Sequence 752, App
c 26	18	0.6	462	10 US-09-864-761-3080	Sequence 3080, Ap
c 27	18	0.6	471	9 US-09-918-995-27364	Sequence 27364, A
c 28	18	0.6	482	9 US-10-040-739-149	Sequence 149, App
c 29	18	0.6	490	9 US-09-918-995-30714	Sequence 30714, A
c 30	18	0.6	496	9 US-09-918-995-11026	Sequence 11026, A
c 31	18	0.6	527	9 US-09-918-995-31632	Sequence 31632, A
c 32	18	0.6	597	10 US-09-864-761-13150	Sequence 13150, A
c 33	18	0.6	682	10 US-09-833-381-836	Sequence 836, App
c 34	18	0.6	954	10 US-09-815-242-9757	Sequence 9757, Ap
c 35	18	0.6	1014	9 US-10-102-806-305	Sequence 305, App
c 36	18	0.6	1376	9 US-10-011-445-18	Sequence 18, Appli
c 37	18	0.6	1410	10 US-09-779-448-3	Sequence 3, Appli
c 38	18	0.6	1536	10 US-09-779-448-4	Sequence 4, Appli
c 39	18	0.6	1859	10 US-09-779-448-2	Sequence 2, Appli
c 40	18	0.6	1953	10 US-09-784-358-9	Sequence 9, Appli
c 41	18	0.6	2175	10 US-09-784-358-7	Sequence 7, Appli
c 42	18	0.6	2316	10 US-09-784-358-13	Sequence 13, Appli
c 43	18	0.6	2457	10 US-09-850-048A-1	Sequence 1, Appli
c 44	18	0.6	2489	10 US-09-764-864-552	Sequence 552, App
c 45	18	0.6	2504	10 US-09-764-864-103	Sequence 103, App

ALIGNMENTS

RESULT 1
US-09-965-631-3
; Sequence 3, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucleotides Enco
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2853
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-965-631-3

Query Match	94.5%	Score 2695;	DB 10;	Length 2853;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 2795;	Conservative	0;	Mismatches	2;
			Indels	0;
			Gaps	0;
QY	57	TGAGCCAGAGCGGAGGTAGTCTGCCATCCGACCTGGACCGGACATTAACGGCGCGG	116	
Db	57	TGAGCCAGAGCGGAGGTAGTCTGCCATCCGACCTGGACCGGACATTAACGGCGCGG	116	
QY	117	CTACTACTGCGGGGTCCCGAGGACTCCGGGGATCAGGGACTATTTTTCAGATCAGC	176	
Db	117	CTACTACTGCGGGGTCCCGAGGACTCCGGGGATCAGGGACTATTTTTCAGATCAGC	176	
QY	177	ATTTCAGGAGACTTTTACCTACCTGAGCGCGGATGCTCAGTCTTGTGCTCCGCCCT	236	
Db	177	ATTTCAGGAGACTTTTACCTACCTGAGCGCGGATGCTCAGTCTTGTGCTCCGCCCT	236	
QY	237	CTCCACTGAGCATCTGGCGCTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGG	296	
Db	237	CTCCACTGAGCATCTGGCGCTCCCTCCAGGGGCTCACCGGGGCTCTTCAGACCTGG	296	
QY	297	ACGCTGCTTCTATTTCTGGGACGCTGAACGCCGAGCGGACTGCTGCTGCTGCTG	356	

|||||
Db 297 ACCTGCTTCTATTTCTGGGACGTGAACGCCGAGCGGACTCGTTTCGCTGCTGAGCCT 356
QY 357 GTCCGGGGGCTCCGCGGAGCCTTTTGGCTACCGAGGCGCGGAGTATGTATTAGCCCGCT 416
Db 357 GTCCGGGGGCTCCGCGGAGCCTTTTGGCTACCGAGGCGCGGAGTATGTATTAGCCCGCT 416
QY 417 GCCCAATGCTAGCGCGCGCGCGGAGCGAGCGACACAGCGCAGGGGCGACACTTCTCCAGCG 476
Db 417 GCCCAATGCTAGCGCGCGCGCGGAGCGAGCGACACAGCGCAGGGGCGACACTTCTCCAGCG 476
QY 477 CCGGGGTGTTCCCGGGGGGCTTCCGAGAGACCCACTCTCGCTCGGGGTGCGCTCGG 536
Db 477 CCGGGGTGTTCCCGGGGGGCTTCCGAGAGACCCACTCTCGCTCGGGGTGCGCTCGG 536
QY 537 CTGGAACCCCGCCACTCTACGGGCCCTTGAACCCCTTACAAGCGCGCGCGGCTTCGG 596
Db 537 CTGGAACCCCGCCACTCTACGGGCCCTTGAACCCCTTACAAGCGCGCGCGGCTTCGG 596
QY 597 GGAGAGTCGTAGCCGGGCGAGGCTTGGGCGCGCCAGCGTTTCGTGCTATCCCGGGTA 656
Db 597 GGAGAGTCGTAGCCGGGCGAGGCTTGGGCGCGCCAGCGTTTCGTGCTATCCCGGGTA 656
QY 657 CGTGGAGACGCTGGTGTGCGGACGAGTCAATGGTCAAGTTCCACGCGCGGACCTGA 716
Db 657 CGTGGAGACGCTGGTGTGCGGACGAGTCAATGGTCAAGTTCCACGCGCGGACCTGA 716
QY 717 ACATTATCTGCTGACGCTGCTGACGCGCGGCGGCGGACTCTACCGCATCCAGCATCCT 776
Db 717 ACATTATCTGCTGACGCTGCTGACGCGCGGCGGCGGACTCTACCGCATCCAGCATCCT 776
QY 777 CAACCCCATCAACATGTTGTGTCAGGTGCTGCTTTCAGAGATCGTACCTCGGGCC 836
Db 777 CAACCCCATCAACATGTTGTGTCAGGTGCTGCTTTCAGAGATCGTACCTCGGGCC 836
QY 837 CAAGGTCACCGGCAATGCGGCCCTGACGCTGCGCAACTTCTGTGCTGCGCAAGAAGCT 896
Db 837 CAAGGTCACCGGCAATGCGGCCCTGACGCTGCGCAACTTCTGTGCTGCGCAAGAAGCT 896
QY 897 GAACAAAGTGTGACAAAGACCCCGAGTACTGGGACACTGCCATGCTTCCACGAGCA 956
Db 897 GAACAAAGTGTGACAAAGACCCCGAGTACTGGGACACTGCCATGCTTCCACGAGCA 956
QY 957 GGACCTGTGTGGAGCACCACCTGTGACACCTTGGGCACTGGGCTGAGTGGGTACCATGTG 1016
Db 957 GGACCTGTGTGGAGCACCACCTGTGACACCTTGGGCACTGGGCTGAGTGGGTACCATGTG 1016
QY 1017 TGACCCCAAGAGAAAGTGTCTGTCATTGAGGACGATGGGCTTCCATCAGCCTTCCACAC 1076
Db 1017 TGACCCCAAGAGAAAGTGTCTGTCATTGAGGACGATGGGCTTCCATCAGCCTTCCACAC 1076
QY 1077 TGCCCAAGAGCTGGGCGACGTTTCAACATGCCCCATGACAATGTGAAGTCTGTGAGGA 1136
Db 1077 TGCCCAAGAGCTGGGCGACGTTTCAACATGCCCCATGACAATGTGAAGTCTGTGAGGA 1136
QY 1137 GGTGTTTGGGAAGCTCCGAGCAACACATGATGTCCCGACCCCTCATCCAGATCGACCG 1196
Db 1137 GGTGTTTGGGAAGCTCCGAGCAACACATGATGTCCCGACCCCTCATCCAGATCGACCG 1196
QY 1197 TGCCAAACCCCTGCTGAGCCTGCTGCCATCATCAGGACTTCTTCCGAGCGGGCA 1256
Db 1197 TGCCAAACCCCTGCTGAGCCTGCTGCCATCATCAGGACTTCTTCCGAGCGGGCA 1256
QY 1257 CGGTGACTGCTTCTGGACCAACCCAGAACCCATCTCCCTGCCCCGAGGATCTGCCGGG 1316
Db 1257 CGGTGACTGCTTCTGGACCAACCCAGAACCCATCTCCCTGCCCCGAGGATCTGCCGGG 1316
QY 1317 CGCCAGCTACACCTTGAGCCAGCTGCGAGCTGCTTTTGGCGTGGGCTCCAGCCCTG 1376
Db 1317 CGCCAGCTACACCTTGAGCCAGCTGCGAGCTGCTTTTGGCGTGGGCTCCAGCCCTG 1376
QY 1377 TCCTTACATGCAAGTACTGACCAAGCTGTGTGTCACCGGAGGCGCAAGGACAGATGGT 1436
|||||

Db 1377 TCCTTACATGCAAGTACTGACCAAGCTGTGTGTCACCGGGAAGGCCAAGGACAGATGGT 1436
QY 1437 GTGCCAGACCCCGCCACTTCCCTTGGGCGGATGACACAGCTGTGCGGAGGCAAGCTCTG 1496
Db 1437 GTGCCAGACCCCGCCACTTCCCTTGGGCGGATGACACAGCTGTGCGGAGGCAAGCTCTG 1496
QY 1497 CCTTAAAGGGGCTGCTGGTGGAGACACAACTTCAACAGCACAGGCTGAGTGTCTCTG 1556
Db 1497 CCTTAAAGGGGCTGCTGGTGGAGACACAACTTCAACAGCACAGGCTGAGTGTCTCTG 1556
QY 1557 GSCCAATGGGATCCCTATGCCCCTGCTCGGCACATGTGTGGGGCGCTGACGCTGGC 1616
Db 1557 GSCCAATGGGATCCCTATGCCCCTGCTCGGCACATGTGTGGGGCGCTGACGCTGGC 1616
QY 1617 CAGGAGGAGTGCACCAACCCCTGSCCAAGGGGGCAAGTACTGCCAGGAGTGA 1676
Db 1617 CAGGAGGAGTGCACCAACCCCTGSCCAAGGGGGCAAGTACTGCCAGGAGTGA 1676
QY 1677 GGTGAAATACCGATCTCTGAACCTTGGAGCCCTGCCCAAGCTCAGCCTCCGGAAGCTT 1736
Db 1677 GGTGAAATACCGATCTCTGAACCTTGGAGCCCTGCCCAAGCTCAGCCTCCGGAAGCTT 1736
QY 1737 CCGGAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGCACCAACCGGCTCACTCT 1796
Db 1737 CCGGAGGAGCAGTGTGAGGCTTTCAACGGCTACAACACAGCACCAACCGGCTCACTCT 1796
QY 1797 CGCCTGTCATGGGTGCCCAAGTACTCCGGGCTGTCTCCCGGGACAAAGTGCAGCTCAT 1856
Db 1797 CGCCTGTCATGGGTGCCCAAGTACTCCGGGCTGTCTCCCGGGACAAAGTGCAGCTCAT 1856
QY 1857 CTGCCGAGCCTAATGGCAGTGGCTACTTCTATGTCTGGCACCCCAAGTGTGGTGGAC 1916
Db 1857 CTGCCGAGCCTAATGGCAGTGGCTACTTCTATGTCTGGCACCCCAAGTGTGGTGGAC 1916
QY 1917 GCTGTGCTCTCTGACTCCACTCCGCTGTGTCCAAGGCAAGTGCATCAAGCTGCTG 1976
Db 1917 GCTGTGCTCTCTGACTCCACTCCGCTGTGTCCAAGGCAAGTGCATCAAGCTGCTG 1976
QY 1977 TGATGGGAACCTGGGCTCCAAAGAGATTCGACAAGTGTGGGCTGTGTGGGGAGACAA 2036
Db 1977 TGATGGGAACCTGGGCTCCAAAGAGATTCGACAAGTGTGGGCTGTGTGGGGAGACAA 2036
QY 2037 TAAGAGCTGCAAGAGTGTGACTGACTTTCACCAAGCCCATGCTATGCTTACATTTCTG 2096
Db 2037 TAAGAGCTGCAAGAGTGTGACTGACTTTCACCAAGCCCATGCTATGCTTACATTTCTG 2096
QY 2097 GGTGCCATCCCGCAGCGGCTCAAGCATCGACATCGCCAGCGCGTTCACAAAGGCT 2156
Db 2097 GGTGCCATCCCGCAGCGGCTCAAGCATCGACATCGCCAGCGCGGTTACAAAGGCT 2156
QY 2157 GATCGGGATGACAACCTACCTGGCTCTGAAGAACAGCAAGCAAGTACCTGCTCAACGG 2216
Db 2157 GATCGGGATGACAACCTACCTGGCTCTGAAGAACAGCAAGCAAGTACCTGCTCAACGG 2216
QY 2217 GCATTTGCTGTGTGTCGGGCTGGAGCGGACCTGGTGGTGAAGGGCAGTCTGCTGGGTA 2276
Db 2217 GCATTTGCTGTGTGTCGGGCTGGAGCGGACCTGGTGGTGAAGGGCAGTCTGCTGGGTA 2276
QY 2277 CAGCGGACGCGGACAGCGGTGGAGAGCCTTCAGGCTTCCCGGCCCATCTCTGCAAGCGCT 2336
Db 2277 CAGCGGACGCGGACAGCGGTGGAGAGCCTTCAGGCTTCCCGGCCCATCTCTGCAAGCGCT 2336
QY 2337 GACCGTGGAGTCTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTCTTA 2396
Db 2337 GACCGTGGAGTCTCTCTCCGTGGGGAAGATGACACCGCCCGGGTCCGCTACTCTCTTA 2396
QY 2397 TCTGCCAAAGAGCTCGGAGGACAGTCTCTCAATCCCAAGGACCCCGGGGACCCCTC 2456
Db 2397 TCTGCCAAAGAGCTCGGAGGACAGTCTCTCAATCCCAAGGACCCCGGGGACCCCTC 2456
QY 2457 TGTCTTGCACAAAGCTCTCTCAGCCTCTCCAAACAGGTCGAGCAGCGGACGACAGGCC 2516
Db 2457 TGTCTTGCACAAAGCTCTCTCAGCCTCTCCAAACAGGTCGAGCAGCGGACGACAGGCC 2516

QY 2517 CCTGACGCTGGTGGCTGGAGCTGGGGCCGCTGCTCCGGAGCTGGCGAGTGGCT 2576
 |||||
 Db 2517 CCTGACGCTGGTGGCTGGAGCTGGGGCCGCTGCTCCGGAGCTGGCGAGTGGCT 2576
 |||||
 QY 2577 GCAGAAAGCGGGGCTGGAGCTGGCGGGCTCCCGGGGACGACAGCTGCTGCTGTA 2636
 |||||
 Db 2577 GCAGAAAGCGGGGCTGGAGCTGGCGGGCTCCCGGGGACGACAGCTGCTGTA 2636
 |||||
 QY 2637 TGACAGCCATCGCCCGTGGAGACAAAGCTGGCGGGAGCCCTGCCACCTGGAGCT 2696
 |||||
 Db 2637 TGACAGCCATCGCCCGTGGAGACAAAGCTGGCGGGAGCCCTGCCACCTGGAGCT 2696
 |||||
 QY 2697 CAGCGCTGGTCAACCTGCTCCAAAGAGCTGGCGGGGATTTAGAGCGCTCACTCAA 2756
 |||||
 Db 2697 CAGCGCTGGTCAACCTGCTCCAAAGAGCTGGCGGGGATTTAGAGCGCTCACTCAA 2756
 |||||
 QY 2757 GTGTGGGCGACGAGCGCGGTGCTGGCGGGACCAAGTCAACTTGCACCGCAAGCC 2816
 |||||
 Db 2757 GTGTGGGCGACGAGCGCGGTGCTGGCGGGACCAAGTCAACTTGCACCGCAAGCC 2816
 |||||
 QY 2817 CCAGGAGCTGGAGCTTCTGCTGCTGAGGCGGCTGTA 2853
 |||||
 Db 2817 CCAGGAGCTGGAGCTTCTGCTGCTGAGGCGGCTGTA 2853
 |||||

RESULT 2

US-09-965-631-7
 ; Sequence 7, Application US/09965631
 ; Patent No. US20020115842A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Friddle, Carl Johan
 ; APPLICANT: Hilbun, Erin
 ; TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucleotides Encodin
 ; FILE REFERENCE: LEX-0241-USA
 ; CURRENT APPLICATION NUMBER: US/09/965,631
 ; PRIOR FILING DATE: 2001-09-27
 ; PRIOR APPLICATION NUMBER: US 60/236,689
 ; PRIOR FILING DATE: 2000-09-29
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 7
 ; LENGTH: 3446
 ; TYPE: DNA
 ; ORGANISM: homo sapiens
 ; US-09-965-631-7

Query Match 94.5%; Score 2695; DB 10; Length 3446;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2795; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 57 TCAGCCAGAGCGGGAGGTAGTCTCCATCGAGCTGGACCGGACATTAACGGCGCGG 116
 |||||
 Db 453 TGAGCCAGAGCGGGAGGTAGTCTCCATCGAGCTGGACCGGACATTAACGGCGCGG 512
 |||||
 QY 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTCAGATCACAGC 176
 |||||
 Db 513 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTCAGATCACAGC 572
 |||||
 QY 177 ATTTTCAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTCTTGGCTCCCGCTT 236
 |||||
 Db 573 ATTTTCAGGAGGACTTTTACCTACACCTGACCGCGGATGCTCAGTCTTGGCTCCCGCTT 632
 |||||
 QY 237 CTCCTACTGAGATCTGGGGCTGCCCTCCAGGGGCTACCGGGGGCTCTTCAGACCTGCG 296
 |||||
 Db 633 CTCCTACTGAGATCTGGGGCTGCCCTCCAGGGGCTACCGGGGGCTCTTCAGACCTGCG 692
 |||||
 QY 297 ACGTGTCTTCTATTCTGGGACGTGAACCGCGAGCGGAGCTGCTGCTGCTGAGCCT 356
 |||||
 Db 693 ACGTGTCTTCTATTCTGGGACGTGAACCGCGAGCGGAGCTGCTGCTGCTGAGCCT 752
 |||||
 QY 357 GTGGGGGGGCTCCGCGGAGCTTTGGCTACCGAGGCGCGGAGTATGTCTATTAGCCGCT 416
 |||||

Db 753 GTGGGGGGGGTCCGCGGAGGCTTTGGCTACCGAGGCGCGAGTATGTCTATTAGCCGCT 812
 |||||
 QY 417 GCCCAATCTAGCGCGCGCGCGCGAGCGCAACAGCGCAGGCGCACACCTTTCTCCAGCG 476
 |||||
 Db 813 GCCCAATCTAGCGCGCGCGCGCGAGCGCAACAGCGCAGGCGCACACCTTTCTCCAGCG 872
 |||||
 QY 477 CGGGGTTTCCGGGGCGGGCTTCCGGAGACCCACACCTCTCGCTGCGGGGGTGGCTCGGG 536
 |||||
 Db 873 CGGGGTTTCCGGGGCGGGCTTCCGGAGACCCACACCTCTCGCTGCGGGGGTGGCTCGGG 932
 |||||
 QY 537 CTGAAACCCGCCCTCTACGGGCGCTTACAGCCCTTACAGCGCGCGGGGCTGGCTTCGG 596
 |||||
 Db 933 CTGNAACCCGCCCTCTACGGGCGCTTACAGCCCTTACAGCGCGCGGGGCTGGCTTCGG 992
 |||||
 QY 597 GGAGAGTGTAGCGCGGCGAGTCTGGGCGCGCAAGCGTTTCTGCTCTATCCCGCGGTA 656
 |||||
 Db 993 GGAGAGTGTAGCGCGGCGAGTCTGGGCGCGCAAGCGTTTCTGCTCTATCCCGCGGTA 1052
 |||||
 QY 657 CGTGAGAGCTGTGTGCGGGGAGTCAAGTCAAGTTCACGGCGCGGAGCTGGA 716
 |||||
 Db 1053 CGTGAGAGCTGTGTGCGGGGAGTCAAGTCAAGTTCACGGCGCGGAGCTGGA 1112
 |||||
 QY 717 ACATTATCTGTGACGCTGTGCAAGCGGGCGCGAGCTCTACCGCCATCTCAGATCCT 776
 |||||
 Db 1113 ACATTATCTGTGACGCTGTGCAAGCGGGCGCGAGCTCTACCGCCATCTCAGATCCT 1172
 |||||
 QY 777 CAACCCCATCAACATCGTTGTGTCAGAGTGTCTCTTAGAGATCGTGA2TCGGGCG 836
 |||||
 Db 1173 CAACCCCATCAACATCGTTGTGTCAGAGTGTCTCTTAGAGATCGTGA2TCGGGCG 1232
 |||||
 QY 837 CAAGGTCAACGGCAATCGGCGCTGACGCTGCGCAACTTCTGTGCGTGGCAGAGAGCT 896
 |||||
 Db 1233 CAAGGTCAACGGCAATCGGCGCTGACGCTGCGCAACTTCTGTGCGTGGCAGAGAGCT 1292
 |||||
 QY 897 GAACAAAGTGTGAGTGAACAGCACCCGAGTACTGGGACACTGCCATCCTCTTACCAGGCA 956
 |||||
 Db 1293 GAACAAAGTGTGAGTGAACAGCACCCGAGTACTGGGACACTGCCATCCTCTTACCAGGCA 1352
 |||||
 QY 957 GGACTGTGTGGAGCCACCACTGTGACACCTTGGGATGGCTGATGTGGTACCATGTG 1016
 |||||
 Db 1353 GGACTGTGTGGAGCCACCACTGTGACACCTTGGGATGGCTGATGTGGTACCATGTG 1412
 |||||
 QY 1017 TGACCCCAAGAGAGTGTCTGTCTATTGAGGAGTGGCTTCCATCAGCCCTTACCAC 1076
 |||||
 Db 1413 TGACCCCAAGAGAGTGTCTGTCTATTGAGGAGTGGCTTCCATCAGCCCTTACCAC 1472
 |||||
 QY 1077 TGCCACAGAGTGGCGCACCTGTTCACATGCCCCCATGACAATGTGAAAGTCTGTGAGGA 1136
 |||||
 Db 1473 TGCCACAGAGTGGCGCACCTGTTCACATGCCCCCATGACAATGTGAAAGTCTGTGAGGA 1532
 |||||
 QY 1137 GGTGTTTGGAAAGTCCGAGCCCAACACATGATGTCCCCGACCTCATCCAGATCGACCG 1196
 |||||
 Db 1533 GGTGTTTGGAAAGTCCGAGCCCAACACATGATGTCCCCGACCTCATCCAGATCGACCG 1592
 |||||
 QY 1197 TGCCAAACCCCTGTGACGCTGAGTGTGCCATCATCAGGACTTCTGTGACAGCGGGA 1256
 |||||
 Db 1593 TGCCAAACCCCTGTGACGCTGAGTGTGCCATCATCAGGACTTCTGTGACAGCGGGA 1652
 |||||
 QY 1257 CGGTGACTGCTCTTGACCAACCCAGCAAGCCATCTCCCTGCGCGAGGATCTCCGGG 1316
 |||||
 Db 1653 CGGTGACTGCTCTTGACCAACCCAGCAAGCCATCTCCCTGCGCGAGGATCTCCGGG 1712
 |||||
 QY 1317 CGCCAGCTACACCTGAGCCAGCAGTGGCTTTTGGCGTGGCTTTTGGCGTGGCTTCCAGGCCCTG 1376
 |||||
 Db 1713 CGCCAGCTACACCTGAGCCAGCAGTGGCTTTTGGCGTGGCTTTTGGCGTGGCTTCCAGGCCCTG 1772
 |||||
 QY 1377 TCCTTACATGCACTACTGCACCAAGCTGTGGTGCACCGGGAAGGCAAGGACAGATGT 1436
 |||||
 Db 1773 TCCTTACATGCACTACTGCACCAAGCTGTGGTGCACCGGGAAGGCAAGGACAGATGT 1832
 |||||
 QY 1437 GTGCCAGAGCCGCTCTCCCTGGCGGATGCCACCACTGTGGGAGGGAAGTCTGTG 1496
 |||||
 Db 1833 GTGCCAGAGCCGCTCTCCCTGGCGGATGCCACCACTGTGGGAGGGAAGTCTGTG 1892
 |||||

```
QY 1497 CCTCAAGGGGCTGGTGGAGAGACAACTACACAGCACAGGGTGGATGTTCTCTG 1556
Db 1893 CCTCAAGGGGCTGGTGGAGAGACAACTACAAAGCACAGGGTGGATGTTCTCTG 1952
QY 1557 GGCCTAATGGATCCCTATATGGCCCTGCTCGGCGACATGTGGTGGGGCGTGCAGCTGGC 1616
Db 1953 GGCCTAATGGATCCCTATATGGCCCTGCTCGGCGACATGTGGTGGGGCGTGCAGCTGGC 2012
QY 1617 CAGGAGCAGTGCACCAACCCCAACCCCTGCGCAACGGGGGCAAGTACTGCGAGGGAGTGAG 1676
Db 2013 CAGGAGCAGTGCACCAACCCCAACCCCTGCGCAACGGGGGCAAGTACTGCGAGGGAGTGAG 2072
QY 1677 GGTGAATACCGATCCTGCAACTCTGAGCCCTGAGCCCTGCGCCAGCTCAGCCCTCGGAAAGAGCTT 1736
Db 2073 GGTGAATACCGATCCTGCAANTCTGAGCCCTGAGCCCTGCGCCAGCTCAGCCCTCGGAAAGAGCTT 2132
QY 1737 CCGGGAGGAGCAGTGTGAGGCTTTTCAACGCTTACACGCTACACACAGCAACCCGCTCACTCT 1796
Db 2133 CCGGGAGGAGCAGTGTGAGGCTTTTCAACGCTTACACGCTACACACAGCAACCCGCTCACTCT 2192
QY 1797 CGCGTGGCATGGTGGCCCAAGTACTCGGCGTGTCTCCCGGGACAAGTCAAGTCTAT 1856
Db 2193 CGCGTGGCATGGTGGCCCAAGTACTCGGCGTGTCTCCCGGGACAAGTCAAGTCTAT 2252
QY 1857 CTGCGAGCCCAATGGCACTGGCTACTTCTATGTCTGGCACCAAGGTGGTGGACGGCAC 1916
Db 2253 CTGCGAGCCCAATGGCACTGGCTACTTCTATGTCTGGCACCAAGGTGGTGGACGGCAC 2312
QY 1917 GCTGTGCTCTCTGACTCCACTCCGCTGTGTGTCTCAAGGCAAGTGCATCAAGCTGGCTG 1976
Db 2313 GCTGTGCTCTCTGACTCCACTCCGCTGTGTGTCTCAAGGCAAGTGCATCAAGCTGGCTG 2372
QY 1977 TGATGGGAACCTGGGCTCCCAAGAGAGATTCGACAAGTGTGGGTGTGGGGGAGACAA 2036
Db 2373 TGATGGGAACCTGGGCTCCCAAGAGAGATTCGACAAGTGTGGGTGTGGGGGAGACAA 2432
QY 2037 TAAGAGCTGCAAGAAGTGAAGTCTTTCACCAAGCCCATGCTGGCTACAAATTTGCT 2096
Db 2433 TAAGAGCTGCAAGAAGTGAAGTCTTTCACCAAGCCCATGCTGGCTACAAATTTGCT 2492
QY 2097 GGTGGCCATCCCGCAGCGGCTTCAGCATCGACATCCGCGCAGCGGTTTACAAAGGCT 2156
Db 2493 GGTGGCCATCCCGCAGCGGCTTCAGCATCGACATCCGCGCAGCGGTTTACAAAGGCT 2552
QY 2157 GATCGGGATGACAACTACCTGGCTCTGAAGAACAGCAGCAAGCAAGTACCTGCTCAACGG 2216
Db 2553 GATCGGGATGACAACTACCTGGCTCTGAAGAACAGCAGCAAGCAAGTACCTGCTCAACGG 2612
QY 2217 GCATTTCTGCTGCTGGCGGTGAGCGGGACCTGGTGGTGAAGGGCAGTCTGTCGGTA 2276
Db 2613 GCATTTCTGCTGCTGGCGGTGAGCGGGACCTGGTGGTGAAGGGCAGTCTGTCGGTA 2672
QY 2277 CAGCGGACGGGACACGGTGGAGAGCCCTGCAGGCTTCCCGGCCATCTCTGGAGCGCT 2336
Db 2673 CAGCGGACGGGACACGGTGGAGAGCCCTGCAGGCTTCCCGGCCATCTCTGGAGCGCT 2732
QY 2337 GACCGTGGAGTCTCTCTCGTGGGGAAGATGACACCGCCCGCGGTCTGCTCTTCTA 2396
Db 2733 GACCGTGGAGTCTCTCTCGTGGGGAAGATGACACCGCCCGCGGTCTGCTCTTCTA 2792
QY 2397 TCTGCCAAGAGGCTCGGAGGAGCAAGTCTCTCATCCCAAGACCCCGGGAGCCCTC 2456
Db 2793 TCTGCCAAGAGGCTCGGAGGAGCAAGTCTCTCATCCCAAGACCCCGGGAGCCCTC 2852
QY 2457 TGTCTTGACAAACAGCTCTCAGCCCTTCCAAACAGGTGGAGCAGCGGACACAGGCG 2516
Db 2853 TGTCTTGACAAACAGCTCTCAGCCCTTCCAAACAGGTGGAGCAGCGGACACAGGCG 2912
QY 2517 CCCTGACGCTGGGTGGTGGCAGCTGGGGCGGTGCTCCGCGAGCTGCGGCGAGTGGCCT 2576
Db 2913 CCCTGACGCTGGGTGGTGGCAGCTGGGGCGGTGCTCCGCGAGCTGCGGCGAGTGGCCT 2972
```

```
QY 2577 GCAGAAGCGGCGGTGGAGTGCCTCGGGGCTCCGCGGGCAGCGACAGGTCCCTGCCTGTGA 2636
Db 2973 GCAGAAGCGGCGGTGGAGTGCCTCGGGGCTCCGCGGGCAGCGACAGGTCCCTGCCTGTGA 3032
QY 2637 TGCAGCCCATCGGCGGCTGGAGACACAAGCTCGGGGAGCCCTGCCCACTGCGGAGCT 2696
Db 3033 TGCAGCCCATCGGCGGCTGGAGACACAAGCTCGGGGAGCCCTGCCCACTGCGGAGCT 3092
QY 2697 CAGGCGCTGTGCACCCCTGCTCAAGAGCTGCGGCCCGGGGATTTTCAGAGGGCTCAGTCAA 2756
Db 3093 CAGGCGCTGTGCACCCCTGCTCAAGAGCTGCGGCCCGGGGATTTTCAGAGGGCTCAGTCAA 3152
QY 2757 GTGTGTGGGCCACAGGAGCGGCTGCTGGGCCCGGGACAGTGCAACTTGCACCGCAAGCC 2816
Db 3153 GTGTGTGGGCCACAGGAGCGGCTGCTGGGCCCGGGACAGTGCAACTTGCACCGCAAGCC 3212
QY 2817 CCAGGAGCTGACATTCCTGCTGCTGAGCGCGTCTGA 2853
Db 3213 CCAGGAGCTGACATTCCTGCTGCTGAGCGCGTCTGA 3249
```

RESULT 3
US-10-163-316-3
; Sequence 3, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MP101-025P1RNM
; CURRENT APPLICATION NUMBER: US/10163,316
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2469)
US-10-163-316-3

Query Match 50.3%; Score 1435; DB 9; Length 2469;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1485; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 57 TCAGCCAGAGCGGGAGGTAGTCTCCATCGACTGGACCCGGACATTAACGGCGCGG 116
Db 57 TCAGCCAGAGCGGGAGGTAGTCTCCATCGACTGGACCCGGACATTAACGGCGCGG 116
QY 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTTCAGTACAGC 176
Db 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGACTCATTTTTCAGTACAGC 176
QY 177 ATTTTACAGGAGACTTTTACCTACCTGACGCGGATGCTAGTCTTGGCTCCGCGCTT 236
Db 177 ATTTTACAGGAGACTTTTACCTACCTGACGCGGATGCTAGTCTTGGCTCCGCGCTT 236
QY 237 CTCCACTGAGCATCTGGGCGTCCCTCCAGGGGCTCACCAGGGGCTCTTCAGACTCGG 296
Db 237 CTCCACTGAGCATCTGGGCGTCCCTCCAGGGGCTCACCAGGGGCTCTTCAGACTCGG 296
QY 297 AGCTGCTTCTATTTCTGGGGAGTGAACCGCCGAGCGGACTGCTGCTGCTGAGCCT 356
Db 297 AGCTGCTTCTATTTCTGGGGAGTGAACCGCCGAGCGGACTGCTGCTGCTGAGCCT 356
QY 357 GTGCGGGGGGCTCCGCGGAGCCTTTGGCTTACCGAGCGCGGAGTATGTCATTAGCCGCT 416
Db 357 GTGCGGGGGGCTCCGCGGAGCCTTTGGCTTACCGAGCGCGGAGTATGTCATTAGCCGCT 416
```

QY 417 GCCCAATGCTAGCGCGCGCGCGAGCGCAAGAGCCAGGCGCACACCTTCTCCAGCG 476
Db 417 GCCCAATGCTAGCGCGCGCGCGAGCGCAAGAGCCAGGCGCACACCTTCTCCAGCG 476
QY 477 CCGGGGTGTTCCGGGCGGGGCTTCCGGAGACCCACACCTCTCGCTGCGGGGTGGCCTCGG 536
Db 477 CCGGGGTGTTCCGGGCGGGGCTTCCGGAGACCCACACCTCTCGCTGCGGGGTGGCCTCGG 536
QY 537 CTGGAACCCCGCCATCTACGGGCGCTTGGACCTTACAGCCGCGGCGCGGCTTCGG 596
Db 537 CTGGAACCCCGCCATCTACGGGCGCTTGGACCTTACAGCCGCGGCGGCTTCGG 596
QY 597 GGAGAGTCGTAGCGGCGCGAGTCTGGGCGCGCAAGCGCTTCTATCCCGGGTA 656
Db 597 GGAGAGTCGTAGCGGCGCGAGTCTGGGCGCGCAAGCGCTTCTATCCCGGGTA 656
QY 657 CGTGAGACGCTGTGTGTCGCGACGAGTCAATGGTCAAGTTCACGCGCGGAGACCTGA 716
Db 657 CGTGAGACGCTGTGTGTCGCGACGAGTCAATGGTCAAGTTCACGCGCGGAGACCTGA 716
QY 717 ACATTATCTGCTGACGCTGTGTCGAACGCGCGGCGACTTACCGGCATCCCGAGCATCT 776
Db 717 ACATTATCTGCTGACGCTGTGTCGAACGCGCGGCGACTTACCGGCATCCCGAGCATCT 776
QY 777 CAACCCCATCAACATCGTTGTGTCGAAGTGTCTTCTTAGAGTCGTGACTCCGGGCC 836
Db 777 CAACCCCATCAACATCGTTGTGTCGAAGTGTCTTCTTAGAGTCGTGACTCCGGGCC 836
QY 837 CAAGGTCACCGGCAATCGCGGCTGACGCTGGCACTTCTGTGCTGGCAGCAAGCT 896
Db 837 CAAGGTCACCGGCAATCGCGGCTGACGCTGGCACTTCTGTGCTGGCAGCAAGCT 896
QY 897 GAACAAGTAGTGACAGACCCCGAGTACTGGGACACTGGCATCTCTTCCACAGGCA 956
Db 897 GAACAAGTAGTGACAGACCCCGAGTACTGGGACACTGGCATCTCTTCCACAGGCA 956
QY 957 GGACCTGTGTGGAGCACCACCTGTGACACCTGGGCGATGGCTGTGGGTACCATGTG 1016
Db 957 GGACCTGTGTGGAGCACCACCTGTGACACCTGGGCGATGGCTGTGGGTACCATGTG 1016
QY 1017 TGACCCCAAGAGAGTGTCTGTCAATTGAGGACATGGGCTTCCATACGCTTCCACAC 1076
Db 1017 TGACCCCAAGAGAGTGTCTGTCAATTGAGGACATGGGCTTCCATACGCTTCCACAC 1076
QY 1077 TGCCACGAGCTGGGCGACGTGTTCAATGCCCAGTGAAGTGAAGTCTGTGAGGA 1136
Db 1077 TGCCACGAGCTGGGCGACGTGTTCAATGCCCAGTGAAGTGAAGTCTGTGAGGA 1136
QY 1137 GGTGTTTGGGAAGCTCCGAGCAACCATGATGTCCCGACCTCATCCAGATCGACCG 1196
Db 1137 GGTGTTTGGGAAGCTCCGAGCAACCATGATGTCCCGACCTCATCCAGATCGACCG 1196
QY 1197 TGCCAAACCCCTGTCAGCTGAGTGTGCTGATCATACCGACTTCTGAGAGCGGCA 1256
Db 1197 TGCCAAACCCCTGTCAGCTGAGTGTGCTGATCATACCGACTTCTGAGAGCGGCA 1256
QY 1257 CGGTGACTGCTTCCGACCAACCCAGCCATCTCCCTGCGCGAGGATCTGCCGG 1316
Db 1257 CGGTGACTGCTTCCGACCAACCCAGCCATCTCCCTGCGCGAGGATCTGCCGG 1316
QY 1317 CGCCAGCTACACCTGAGCGACGAGTGGAGTGGCTTGGCGTGGCTCCAGCCCTG 1376
Db 1317 CGCCAGCTACACCTGAGCGACGAGTGGAGTGGCTTGGCGTGGCTCCAGCCCTG 1376
QY 1377 TCCTTACATGCACTGACCAAGCTGTGTGACCGGAGGCAAGGACAGATGCT 1436
Db 1377 TCCTTACATGCACTGACCAAGCTGTGTGACCGGAGGCAAGGACAGATGCT 1436
QY 1437 GTGCCAGACCCGCACTTCCCTGCGGCGATGGCAGCAGCTGTGCGAGGCAAGCTCTG 1496
Db 1437 GTGCCAGACCCGCACTTCCCTGCGGCGATGGCAGCAGCTGTGCGAGGCAAGCTCTG 1496
QY 1497 CCTAAAGGGGCTGCTGGTGGAGAGACAAACCTCAAGACACAG 1542

Db 1497 CCTAAAGGGGCTGCTGGTGGAGAGACAAACCTCAAGACACAG 1542

RESULT 4

US-10-163-316-1
; Sequence 1, Application US/10163316
; Publication No. US2002019703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; FILE REFERENCE: MPI01-025P1RNM
; CURRENT APPLICATION NUMBER: US/10163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2940
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (472)...(2941)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2940)
; OTHER INFORMATION: n = A,T,C or G
US-10-163-316-1

Query Match 50.3%; Score 1435; DB 9; Length 2940;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1485; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 57 TGAGCCAGAGCGGGAGGTAGTCGTTCCCATCCGACTCGGACCGCGACATTAACGCGCGCG 116
Db 528 TGAGCCAGAGCGGGAGGTAGTCGTTCCCATCCGACTCGGACCGCGACATTAACGCGCGCG 587
QY 117 CTACTACTGGCGGGTCCCGAGGACTCGGGGATCAGGACTATTTTCAATCACAGC 176
Db 588 CTACTACTGGCGGGTCCCGAGGACTCGGGGATCAGGACTATTTTCAATCACAGC 647
QY 177 ATTTACAGGAGACTTTTACCTACACCTGACGCGGAGTCTCAGTCTTGCTCCCGCTT 236
Db 648 ATTTACAGGAGACTTTTACCTACACCTGACGCGGAGTCTCAGTCTTGCTCCCGCTT 707
QY 237 CTCACATGAGCATCTGGGCGCTCCCGCTCCAGGGGCTCACCGGGGCTCTTACAGACTGCG 296
Db 708 CTCACATGAGCATCTGGGCGCTCCCGCTCCAGGGGCTCACCGGGGCTCTTACAGACTGCG 767
QY 297 ACGTGTCTTCTATTTCTGGGACGTGAACGCGGAGCTCGTTCCTGCTGTGAGCCT 356
Db 768 ACGTGTCTTCTATTTCTGGGCGCTGAACGCGGAGCTCGTTCCTGCTGTGAGCCT 827
QY 357 GTGCGGGGGCTCCCGCGAGGCTTTGGCTACCGAGGCGCGAGTATGCTATTAGCCGCT 416
Db 828 GTGCGGGGGCTCCCGCGAGGCTTTGGCTACCGAGGCGCGAGTATGCTATTAGCCGCT 887
QY 417 GCCCAATGCTAGCGCGCGCGGCGAGCGCAACAGCAGGCGCGACACCTTCTCCAGCG 476
Db 888 GCCCAATGCTAGCGCGCGCGGCGAGCGCAACAGCAGGCGCGACACCTTCTCCAGCG 947
QY 477 CCGGGGTGTTCCGGGCGGGCTTCCGGAGACCCACCTCTCGCTGCGGGGTGGCCTCGG 536
Db 948 CCGGGGTGTTCCGGGCGGGCTTCCGGAGACCCACCTCTCGCTGCGGGGTGGCCTCGG 1007
QY 537 CTGGAACCCCGCCATCTACGGGCGCTTGGACCTTACAGCCGCGGCGGCTTCGG 596
Db 1008 CTGGAACCCCGCCATCTACGGGCGCTTGGACCTTACAGCCGCGGCGGCTTCGG 1067
QY 597 GGAGAGTCGTAGCGGCGCGAGTCTGGGCGCGCAAGCGTTCGTGCTATCCCGGGTA 656

```

1068 GGAGAGTCGTAGCCGGCGCAGGTCTGGCGCGCAAGCGTTTGGTCTATCCCGGGTA 1127
657 CGTGGAGACGCTGGTGGTCCGGACGAGTCAATGGTCAAGTTCCACGGCGGACCTGGA 716
1128 CGTGGAGACGCTGGTGGTCCGGACGAGTCAATGGTCAAGTTCCACGGCGGACCTGGA 1187
717 ACATTATCTGCTGACGCTGTGTCGACGCGCGGCGGACCTCTACCGCATCCGAGCATCT 776
1188 ACATTATCTGCTGACGCTGTGTCGACGCGCGGCGGACCTCTACCGCATCCGAGCATCT 1247
777 CAACCCCATCAACATCTGTTGGTCAAGTGTCTCTTTAGAGATCGTGACTCCGGGCG 836
1248 CAACCCCATCAACATCTGTTGGTCAAGTGTCTCTTTAGAGATCGTGACTCCGGGCG 1307
837 CAAGGTCAACCGGCAATCGGCGCTGACGCTGGCGACACTCTGTGCTTGGCGAGAGCT 896
1308 CAAGGTCAACCGGCAATCGGCGCTGACGCTGGCGACACTCTGTGCTTGGCGAGAGCT 1367
897 GAACAAAGTGAAGTGAACAGCAGCCGAGTACTGGGACACTGCCATCTCTTCAACAGGA 956
1368 GAACAAAGTGAAGTGAACAGCAGCCGAGTACTGGGACACTGCCATCTCTTCAACAGGA 1427
957 GGACCTGTGTGGAGCCACCACTGTGTACACCTTGGGATGGCTGTGTGGTACCATGTG 1016
1428 GGACCTGTGTGGAGCCACCACTGTGTACACCTTGGGATGGCTGTGTGGTACCATGTG 1487
1017 TGACCCCAAGAGAGCTGTCTGTCAATGAGGAGGATGGCTTCCATCAGCTTCAACCAC 1076
1488 TGACCCCAAGAGAGCTGTCTGTCAATGAGGAGGATGGCTTCCATCAGCTTCAACCAC 1547
1077 TGCCACAGAGCTGGGCGGACGCTGTGTCAATGAGGAGGATGGCTTCCATCAGCTTCA 1136
1548 TGCCACAGAGCTGGGCGGACGCTGTGTCAATGAGGAGGATGGCTTCCATCAGCTTCA 1607
1137 GGTGTTTGGGAGCTCCGAGGCAACACATGATGTCCCGGACCTCATCCAGATCGACCG 1196
1608 GGTGTTTGGGAGCTCCGAGGCAACACATGATGTCCCGGACCTCATCCAGATCGACCG 1667
1197 TGCCAAACCCCTGTGTGAGCTGTGTGTGATGATGATGATGATGATGATGATGATGAT 1256
1668 TGCCAAACCCCTGTGTGAGCTGTGTGTGATGATGATGATGATGATGATGATGATGAT 1727
1257 CGGTGACTGCTCTGTGACCAACCCAGACGATGATGATGATGATGATGATGATGATGAT 1316
1728 CGGTGACTGCTCTGTGACCAACCCAGACGATGATGATGATGATGATGATGATGATGAT 1787
1317 CGCAGCTACACCTGTGACGAGCAGTGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1376
1788 CGCAGCTACACCTGTGACGAGCAGTGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1847
1377 TCCTTACATGAGTACTGACCAAGCTGTGTGTGACCGGAGGAGGAGGAGGAGGAGGAGT 1436
1848 TCCTTACATGAGTACTGACCAAGCTGTGTGTGACCGGAGGAGGAGGAGGAGGAGGAGT 1907
1437 GTGCCAGACCGGACCTTCCCTGTGGCGGATGGACGAGTGTGGCGAGGAGGAGGAGT 1496
1908 GTGCCAGACCGGACCTTCCCTGTGGCGGATGGACGAGTGTGGCGAGGAGGAGGAGT 1967
1497 CCTAAAGGGGCGCTGCTGGAGAGACACAACTCTCAAGACACAGG 1542
1968 CCTAAAGGGGCGCTGCTGGAGAGACACAACTCTCAAGACACAGG 2013

```

RESULT 5
 US-09-965-631-5
 ; Sequence 5, Application US/09965631
 ; Patent No. US20020115842A1
 ; GENERAL INFORMATION: Carl Johan
 ; APPLICANT: Friddle, Carl Johan
 ; TITLE OF INVENTION: No. US20020115842A1 Human Proteases and Polynucleotides Encodin
 ; FILE REFERENCE: LEX-0241-USA

```

; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: homo sapiens
; US-09-965-631-5

Query Match      36.3%; Score 1035; DB 10; Length 1104;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1035; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 TGAGCCAGAGCGGGAGGTAGTCTCCATCGACTGACCGGACATTAACGGCGCGCG 116
DB 57 TGAGCCAGAGCGGGAGGTAGTCTCCATCGACTGACCGGACATTAACGGCGCGCG 116
QY 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGTCAATTTTTCAGATCACAG 176
DB 117 CTACTACTGGCGGGTCCCGAGGACTCCGGGATCAGGAGTCAATTTTTCAGATCACAG 176
QY 177 ATTTTCAGAGGAGTCTTTTACCTACACTGACCCGGATGCTCAGTTCTTGGCTCC-SCCTT 236
DB 177 ATTTTCAGAGGAGTCTTTTACCTACACTGACCCGGATGCTCAGTTCTTGGCTCC-SCCTT 236
QY 237 CTCCACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTCCACGGGGGCTCTTCAGACCTCGG 296
DB 237 CTCCACTGAGCATCTGGGCGTCCCGCTCCAGGGGCTCCACGGGGGCTCTTCAGACCTCGG 296
QY 297 AGCTGCTCTTATCTTGGGGAGCTGAACCGCCGAGCGGACTCGTTTCGTGTGAGGCT 356
DB 297 AGCTGCTCTTATCTTGGGGAGCTGAACCGCCGAGCGGACTCGTTTCGTGTGAGGCT 356
QY 357 GTGCGGGGGGCTCCCGGAGCCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCCGCT 416
DB 357 GTGCGGGGGGCTCCCGGAGCCTTTGGCTACCGAGGCGCGAGTATGTCATTAGCCCGCT 416
QY 417 GCCCAATGCTAGCGCGCGCGGCGAGCGCAACAGCAGGCGGACACTTCTCCAGCG 476
DB 417 GCCCAATGCTAGCGCGCGCGGCGAGCGCAACAGCAGGCGGACACTTCTCCAGCG 476
QY 477 CGGGGGTGTTCGGGCGGCGGCTTCGGGAGACCCACTCTCGCTCGGGGTGCGCTCGGG 536
DB 477 CGGGGGTGTTCGGGCGGCGGCTTCGGGAGACCCACTCTCGCTCGGGGTGCGCTCGGG 536
QY 537 CTGGAACCCCGCCATCTTACGGGCGCTTGAGACCTTTACAAGCCGCGGCGGGGCTTCGG 596
DB 537 CTGGAACCCCGCCATCTTACGGGCGCTTGAGACCTTTACAAGCCGCGGCGGGGCTTCGG 596
QY 597 GGAGAGTCTGAGCGCGGAGGTCTGGGGCGGCGCAAGCGTTTTCGTGTCTATCCCGGGTA 656
DB 597 GGAGAGTCTGAGCGCGGAGGTCTGGGGCGGCGCAAGCGTTTTCGTGTCTATCCCGGGTA 656
QY 657 CGTGGAGAGCTGTGTGCTCGGAGAGTCAATGTTCAAGTTCCACGGCGGAGCTGGA 716
DB 657 CGTGGAGAGCTGTGTGCTCGGAGAGTCAATGTTCAAGTTCCACGGCGGAGCTGGA 716
QY 717 ACATTATCTGTGACGCTGCTGGCAACGGCGGCGGACTCTACCGCATCCGAGATCCT 776
DB 717 ACATTATCTGTGACGCTGCTGGCAACGGCGGCGGACTCTACCGCATCCGAGATCCT 776
QY 777 CAACCCCATCAACATCTGTTGGTCAAGTGTCTTCTTAGAGATCGTGACTCCGGGCG 836
DB 777 CAACCCCATCAACATCTGTTGGTCAAGTGTCTTCTTAGAGATCGTGACTCCGGGCG 836
QY 837 CAAGGTCAACCGGCAATGCGGCGCTGACGCTGGGCAACTTCTGTGCTGCGAGAGT 896
DB 837 CAAGGTCAACCGGCAATGCGGCGCTGACGCTGGGCAACTTCTGTGCTGCGAGAGT 896
QY 897 GAACAAAGTGAAGTGAACAGCAGCCGAGTACTGGGACACTGCCATCTCTTCAACAGGA 956

```

Query Match	31.7%	Score 903:	DB 10:	Length 966;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 903;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	57	TGAGCCAGAGCGGGAGGTAGTCGTTCCATCCGACTGGACCGGACATTAAGCGCGCGG	116	
Db	57	TGAGCCAGAGCGGGAGGTAGTCGTTCCATCCGACTGGACCGGACATTAAGCGCGCGG	116	
QY	117	CTACTACTGGCGGGTCCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCACAG	176	
Db	117	CTACTACTGGCGGGTCCGAGGACTCCGGGATCAGGACTCATTTTTCAGATCACAG	176	
QY	177	ATTTACGAGGAGCTTTTACCTACACCTGACCCGGATGCTCAGTTCTTGCTCCCGCCTT	236	
Db	177	ATTTACGAGGAGCTTTTACCTACACCTGACCCGGATGCTCAGTTCTTGCTCCCGCCTT	236	
QY	237	CTCCACTGAGCATCTGGGGGTCCCTCCAGGGGCTCACCGGGGGTCTTTCAGACTGGG	296	
Db	237	CTCCACTGAGCATCTGGGGGTCCCTCCAGGGGCTCACCGGGGGTCTTTCAGACTGGG	296	
QY	297	ACGCTGCTTCTATTCTGGGGAGCTGAACCGGACCGGACTCGCTTCGCTGTGAGCCT	356	
Db	297	ACGCTGCTTCTATTCTGGGGAGCTGAACCGGACCGGACTCGCTTCGCTGTGAGCCT	356	
QY	357	GTGCGGGGGGCTCCGCGGAGCCTTTGGCTACCGAGGCGCGAGTATGTCAATAGCCCGCT	416	
Db	357	GTGCGGGGGGCTCCGCGGAGCCTTTGGCTACCGAGGCGCGAGTATGTCAATAGCCCGCT	416	
QY	417	GCCCAATGCTAGCGCGCGGGCGGACGGCAACAGCCAGGGCGCACACCTTCTCCAGCG	476	
Db	417	GCCCAATGCTAGCGCGCGGGCGGACGGCAACAGCCAGGGCGCACACCTTCTCCAGCG	476	
QY	477	CCGGGGTGTTCGGGGCGGGCCTTCGCGAGACCCACCTCTCGCTGGGGGTGCCTTCGGG	536	
Db	477	CCGGGGTGTTCGGGGCGGGCCTTCGCGAGACCCACCTCTCGCTGGGGGTGCCTTCGGG	536	


```

; Patent No. US20020157132A1
; GENERAL INFORMATION:
; APPLICANT: Falco, S. Carl
; TITLE OF INVENTION: Plant Amino Acid Biosynthetic Enzymes
; FILE REFERENCE: BB1116 US CIP
; CURRENT APPLICATION NUMBER: US/09/931,457A
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: 09/424,976
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/065,385
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/049,406
; PRIOR FILING DATE: 1997-06-12
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; TYPE: DNA
; ORGANISM: Zea mays
US-09-931-457A-10

Query Match      0.8%; Score 22; DB 9; Length 1813;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1659 GTACTGCGAGGAGTGAGGCTG 1680
Db 269 GTACTGCGAGGAGTGAGGCTG 290

RESULT 9
US-09-918-995-22019/c
; Sequence 22019, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 05/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22019
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(506)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-22019

Query Match      0.7%; Score 21; DB 9; Length 506;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1608 GCAGCTGCCAGGAGGCAGTG 1628
Db 346 GCAGCTGCCAGGAGGCAGTG 326

RESULT 10
US-09-925-297-180/c
; Sequence 180, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 180
; LENGTH: 1827
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1524)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-180

Query Match      0.7%; Score 21; DB 10; Length 1827;
Best Local Similarity 100.0%; Pred. No. 0.91;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1608 GCAGCTGCCAGGAGGCAGTG 1628
Db 1606 GCAGCTGCCAGGAGGCAGTG 1586

RESULT 11
US-09-938-842A-60
; Sequence 60, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 60
; LENGTH: 1017
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-60

Query Match      0.7%; Score 20; DB 9; Length 1017;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1996 AAGAAGAGATTGCAAGTG 2015
Db 468 AAGAAGAGATTGCAAGTG 487

RESULT 12
US-09-535-459-1835
; Sequence 1835, Application US/09535459
; Publication No. US20030040615A1
; GENERAL INFORMATION:
; APPLICANT: Seilhamer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
```



```
; APPLICANT: Naughton, Rebecca E.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING OR REGULATING ELECTRON TRANSFER MOLECULE
; FILE REFERENCE: PD-1014 CIP
; CURRENT APPLICATION NUMBER: US/09/535,459
; CURRENT FILING DATE: 2000-03-24
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2170
; SOFTWARE: PERL Program
; SEQ ID NO 1835
; LENGTH: 256
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID NO. US20030040615A1 rat00012080
; NAME/KEY: unsure
; LOCATION: 205-206, 221
; OTHER INFORMATION: a, t, c, g, or other
US-09-535-459-1835

Query Match          0.7%; Score 19; DB 9; Length 256;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 889 AAGAAAGCTCAACAAAGTGA 907
    |||||
Db 238 AAGAAAGCTCAACAAAGTGA 256

RESULT 13
US-09-918-995-15962/C
; Sequence 15962, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15962
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-15962

Query Match          0.7%; Score 19; DB 9; Length 408;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2678 CCTGCCCCACCTGGGAGCT 2696
    |||||
Db 287 CCTGCCCCACCTGGGAGCT 269

RESULT 14
US-09-925-297-85
; Sequence 85, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
```

```
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
; LENGTH: 431
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (325)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (334)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (347)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (381)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-85

Query Match          0.7%; Score 19; DB 10; Length 431;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 829 TCGGGCCCCAAGGTACCG 847
    |||||
Db 44 TCGGGCCCCAAGGTACCG 62

RESULT 15
US-09-918-995-32286
; Sequence 32286, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32286
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(452)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-32286

Query Match          0.7%; Score 19; DB 9; Length 452;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1709 GCCCCAGCTCAGCTCCGG 1727
    |||||
Db 384 GCCCCAGCTCAGCTCCGG 402

Search completed: May 2, 2003, 01:19:03
Job time : 382 secs
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 01:01:17 ; Search time 3889 seconds
(without alignments)
11881.132 Million cell updates/sec

Title: US-10-009-332-2

Perfect score: 2853
Sequence: 1 atgctttgtggcattct.....gcgtctgagccgtgctga 2853

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estlin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vit:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	53	1.9	544	12 BF078689	BF078689 229090 MA
c 2	39	1.4	195	10 AW935807	AW935807 QV3-DT001
c 3	38	1.3	528	17 A0809642	A0809642 HS.4758.A
c 4	36	1.3	961	10 B8612189	B8612189 B8612189
c 5	35	1.2	469	17 A2842795	A2842795 2M0141119
c 6	35	1.2	522	10 BE553572	BE553572 ur47b04.y

7	1.1	305	12	BF320986	BF320986 uz57hl0.y
8	1.0	400	17	AQ440250	AQ440250 HS_5071.B
9	1.0	415	17	AW485848	AW485848 68315 MAR
10	27	169	17	B01986	B01986 CSRL-145A3-
11	26	627	10	BE281680	BE281680 601099657
12	25	247	10	BB604671	BB604671 BB604671
13	25	849	13	BL103177	BL103177 602889417
14	24	464	10	BE666088	BE666088 149119 MA
15	24	664	13	BF692986	BF692986 603344521
16	23	681	13	BG921734	BG921734 602825569
17	23	1011	17	CNS0471P	AL277414 Tetraodon
18	22	281	10	BB195208	BB195208 BB195208
19	22	314	10	BB154064	BB154064 BB154064
20	22	577	9	AI770556	AI770556 606054C06
21	22	614	14	B0037780	B0037780 946142B02
22	22	649	10	AW061788	AW061788 660012A07
23	22	2737	11	AY109369	AY109369 Zea mays
24	21	109	9	AI423690	AI423690 tf85f08.x
25	21	123	17	CNS010USU	AL168231 Tetraodon
26	21	207	10	AW105581	AW105581 xg33h10.x
27	21	215	9	AA303391	AA303391 EST15998
28	21	221	9	AA599823	AA599823 ag09b04.s
29	21	233	10	BB075309	BB075309 qu83c05.x
30	21	235	10	BB075309	BB075309 BB075309
31	21	238	10	AW046359	AW046359 UI-M-BH1-
32	21	241	9	AI233548	AI233548 EST230236
33	21	244	9	AI449663	AI449663 mr52e05.x
34	21	245	10	AW130644	AW130644 xg64a07.x
35	21	255	9	AI202856	AI202856 q15ahl1.x
36	21	256	12	BF283275	BF283275 EST447866
37	21	256	13	BI534593	BI534593 397660 MA
38	21	260	9	AI698894	AI698894 wc95f04.x
39	21	261	10	AW430043	AW430043 68812 MAR
40	21	278	9	AA996648	AA996648 UI-R-CO-h
41	21	283	12	BF511038	BF511038 UI-H-B14-
42	21	301	9	AA478616	AA478616 zv19c12.s
43	21	302	9	AI499824	AI499824 tm92h10.x
44	21	305	9	AI580450	AI580450 to34a06.x
45	21	311	10	BE645522	BE645522 7e19i08.x

ALIGNMENTS

RESULT 1
BF078689/c
LOCUS 544 bp mRNA linear EST 18-OCT-2000
DEFINITION 229090 MARC 2PTG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF078689
VERSION BF078689.1 GI:10872519
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 544)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.F.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.H., Laegreid, W.W.
and Keeler, J.W.

REFERENCE
AUTHORS
TITLE Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -mnscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT

BACKWARD: GTTTCACGTCAGCAGC
Plate: 43 row: P column: 17
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..544
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2Pig"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 117 a 154 c 169 g 104 t

ORIGIN

Query Match 1.9%; Score 53; DB 12; Length 544;
Best Local Similarity 100.0%; Pred. No. 5.1e-14;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 742 ACGCGCGCGGACTCTACCGCCATCCAGCATCTCAACGCCATCAACATCGT 794
|||||
Db 361 ACGCGCGCGGACTCTACCGCCATCCAGCATCTCAACGCCATCAACATCGT 309
|||||

RESULT 2
AW935807/C
LOCUS AW935807 195 bp mRNA linear EST 30-MAY-2000
DEFINITION QV3-DT0019-081299-039-c08 DT0019 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW935807
VERSION AW935807.1 GI:8111213
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 195)
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-QV3-DT0019-081
299-039-c08&t3=1999-12-08&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 58
High quality sequence stop: 109.
Location/Qualifiers
1..195
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="DT0019"
/dev_stage="Adult"
/note="Organ: denis;drash; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 37 a 60 c 52 g 46 t

ORIGIN

Query Match 1.4%; Score 39; DB 10; Length 195;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1542 GGTGATGTTCTTCGGGCCAAATGGATCCCTATGGCCC 1580
|||||
Db 155 GGTGATGTTCTTCGGGCCAAATGGATCCCTATGGCCC 117
|||||

RESULT 3
AQ809642
LOCUS AQ809642 528 bp DNA linear GSS 10-AUG-1999
DEFINITION HS_4758_Al_D11_77A CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=4758 Col=21 Row=G, DNA sequence.
ACCESSION AQ809642
VERSION AQ809642.1 GI:5728884
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 528)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 4758 row: G column: 21
Seq primer: T7
Class: BAC ends
High quality sequence stop: 528.
Location/Qualifiers
1..528
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=4758 Col=21 Row=G"
/note="male"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 138 a 145 c 100 g 142 t 3 others

ORIGIN

Query Match 1.3%; Score 38; DB 17; Length 528;
Best Local Similarity 100.0%; Pred. No. 7.9e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2114 GCGCCTCAAGCATCGACATCCGCCAGCGGTTACAAA 2151
|||||
Db 296 GCGCCTCAAGCATCGACATCCGCCAGCGGTTACAAA 333
|||||

RESULT 4
BB612189
LOCUS BB612189 961 bp mRNA linear EST 26-OCT-2001
DEFINITION BB612189 RIKEN full-length enriched, 0 day neonate skin Mus
musculus cDNA clone 4631401M01 5', mRNA sequence.

BB612189
 BB612189.1 GI:16453236
 EST.
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 961)
 Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
 Hiramoto, K., Ishii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda,
 M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Onno, M.,
 Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki,
 D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
 Tagami, M., Tagawa, A., Takanashi, F., Takeda, Y., Tanaka, T., Toya, T.,
 Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
 UNPUBLISHED (2001)
 CONTACT: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gs.riken.go.jp/
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh,
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura,
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara,
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa,
 K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
 Hayashizaki, Y.
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001).
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.
 e mouse tissues.
 Location/Qualifiers
 1..961
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="4631401M01"
 /clone_lib="RIKEN full-length enriched, 0 day neonate
 skin"
 /sex="mixed"
 /tissue_type="skin"
 /dev_stage="0 day neonate"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by

cap-trapper. cDNA went through one round of normalization
 to Rot = 10.0 and subtraction to Rot = 100.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence 15' GAGAGAGATTCGATTAAATTAATCCCCCCCCCC
 3'. cDNA was cloned into the XhoI and BamHI sites.
 Vector: a modified pBluescript KS(+) after bulk excision
 from Lambda FLC I"
 BASE COUNT 158 a 302 c 200 t
 ORIGIN
 Query Match 1.3% Score 36; DB 10; Length 961;
 Best Local Similarity 100.0%; Pred. No. 8.3e+06;
 Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 449 ACAGCCAGGGCGCACACCTTCTCCAGCGCGGGGTG 484
 ||||||||||||||||||||||||||||||||||||
 Db 768 ACAGCCAGGGCGCACACCTTCTCCAGCGCGGGGTG 803
 ||||||||||||||||||||||||||||||||||||
 RESULT 5
 A2842795/c
 LOCUS
 DEFINITION
 2M0141119f Mouse 10kb plasmid UUGCIM library Mus.musculus genomic
 clone UUGC2M014119 F, DNA sequence.
 ACCESSION
 VERSION
 A2842795
 A2842795.1 GI:13012703
 GSS
 house mouse.
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 469)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.,
 and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 UNPUBLISHED (2000)
 CONTACT: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: rdunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0141 row: 1 column: 19
 Seq primer: CGTTGTAACGACGCGCCAGT
 Class: plasmid ends
 High quality sequence stop: 469.
 Location/Qualifiers
 1..469
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M014119"
 /clone_lib="Mouse 10kb plasmid UUGCIM library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PWD42uv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (<http://www.jax.org/resources/documents/dnares/>). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gil14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 122 a 151 c 87 g 109 t
ORIGIN

Query Match 1.2%; Score 35; DB 17; Length 469;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1942 GTCTGTGTCACCAAGTCATCAAGCTGGCTG 1976
|||||

Db 327 GTCTGTGTCACCAAGTCATCAAGCTGGCTG 293
|||||

RESULT 6
BE553572 522 bp mRNA linear EST 15-AUG-2000
LOCUS ur47b04.y1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:3153391 5'
DEFINITION similar to TR:054768 054768 A DISINTEGRIN AND METALLOPROTEINASE WITH
THROMBOSPONDIN MOTIFS ; mRNA sequence.

ACCESSION BE553572 1 GI:9817989

VERSION BE553572

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;

TITLE 1 (bases 1 to 522)

JOURNAL NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP);
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1056147

Seq primer: -40RP from Gibco

High quality sequence stop: 385.

Location/Qualifiers

1..522

/organism="Mus musculus"

/strain="FVB/N-3"

/db_xref="taxon:10090"

/clone="IMAGE:3153391"

/clone_lib="NCI_CGAP_Mam2"

/tissue_type="tumor, biopsy sample"

/dev_stage="5 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;

Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Library constructed by Life Technologies. Investigator

providing samples: Gilbert Smith, NIH"

BASE COUNT 116 a 133 c 158 g 115 t

ORIGIN

Query Match 1.2%; Score 35; DB 10; Length 522;

Best Local Similarity 100.0%; Pred. No. 2.2e-05;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DEFINITION uz57h10.y1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3673219 5'
similar to TR:Q9UP80 Q9UP80 METHI PROTEIN. ; mRNA sequence.
ACCESSION BF320986
VERSION BF320986.1 GI:11270026
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;

TITLE 1 (bases 1 to 305)

JOURNAL NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP);
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml

MG1:1433987

High quality sequence stop: 192.

Location/Qualifiers

1..305

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:3673219"

/clone_lib="NCI_CGAP_Mam6"

/sex="female, virgin"

/tissue_type="infiltrating ductal carcinoma"

/dev_stage="5 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;

Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Library constructed by Life Technologies. Investigator

providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 79 a 71 c 85 g 69 t

ORIGIN

Query Match 1.1%; Score 30; DB 12; Length 305;

Best Local Similarity 100.0%; Pred. No. 0.0048;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1765 GGCTACCAACACAGCACCACCGGCTCACT 1794
|||||

Db 48 GGCTACCAACACAGCACCACCGGCTCACT 77
|||||

RESULT 8

AQ440250 400 bp DNA linear GSS 31-MAR-1999

LOCUS HS_5071.B2.D07.T7A RPCI-11 Human Male BAC Library Hmo sapiens

DEFINITION genomic clone Plate-647 Col-14 Row-H, DNA sequence.

ACCESSION AQ440250

VERSION AQ440250.1 GI:4551589

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 400)

JOURNAL Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

COMMENT Koller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hoed,L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

99380589

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering.bac.htm>) or from Research Genetics (<http://info@resgen.com>). BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 647 row: H column: 14
Seq primer: T7
Class: BAC ends
High quality sequence stop: 400.

FEATURES
source

Location/Qualifiers
1. .400
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate:647 Col=14 Row=H"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT 100 a 116 c 98 g 83 t 3 others
ORIGIN

Query Match 1.0%; Score 29; DB 17; Length 400;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2164 GATGCAACTACTCGCTCTGAAGACAG 2192
|||||
DB 359 GATGCAACTACTCGCTCTGAAGACAG 387

RESULT 9

AW485848
LOCUS 415 bp mRNA linear EST 09-JUL-2000
DEFINITION 68315 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION AW485848
VERSION AW485848.1 GI:7055954
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 415)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keeler, J.W.
Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine

Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 20 and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 28 row: D column: 18
Seq primer: ATTTAGGTGACACTATAG.

FEATURES
source

Location/Qualifiers
1. .415
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 70 a 141 c 141 g 63 t
ORIGIN

Query Match 1.0%; Score 28; DB 10; Length 415;
Best Local Similarity 100.0%; Pred. No. 0.047;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2808 CCGCAAGCCCCAGGAGCTGGACTTCTGC 2835
|||||
DB 35 CCGCAAGCCCCAGGAGCTGGACTTCTGC 62

RESULT 10

B01986
LOCUS 169 bp DNA linear GSS 13-JUL-1996
DEFINITION CSRL-145A3-u CSRL flow sorted Chromosome 11 specific cosmid Homo sapiens genomic clone CSRL-145A3, DNA sequence.
ACCESSION B01986
VERSION B01986.1 GI:14111264
KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 169)
Evans, G.A., Burbee, D., Davies, C., Hahner, L., Oliver, T., Gilbert, M., Jones, D., Ward, T., Gillilan, E., Schagemann, J., Probst, S., Harris, J., DeFord, J., McFarland, J., Burzinski, K., Khan, M., Kupfer, K. and Garner, H.R.

TITLE
JOURNAL
COMMENT

Genomic Sequence Sampled Map of Chromosome 11
Unpublished (1996)
Contact: Evans GA, Shane Probst
Mcdermott Center for Human Growth and Development
University of Texas Southwestern Medical Center At Dallas
5323 Harry Hines Blvd, Dallas TX 75235-8591
Tel: 214-648-1600
Fax: 214-648-1666
Email: gevanseutsw.swmed.edu, shane@mcdermott.swmed.edu
Seq primer: T7
Class: cosmid ends
High quality sequence stop: 169.

FEATURES

source
Location/Qualifiers
1. .169
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSRL-145A3"
/clone_lib="CSRL flow sorted Chromosome 11 specific cosmid"
/sex="female"
/cell_type="chimeric hamster somatic cell hybrid"
/note="Vector: sCos-1; Human Chromosome 11 specific cosmid library prepared from flow sorted human Chromosome 11 derived from Chinese Hamster Ovary (CHO) monochromosomal somatic cell hybrid, J1"

BASE COUNT 26 a 56 c 41 g 36 t 10 others
ORIGIN

Query Match 0.9%; Score 27; DB 17; Length 169;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 475 CGCCGGGGTTCGCGGGCGGCTTCC 501
|||||

```

Db 56 CGCGGGGTGTCGGGGGGGCGCTTCC 82
RESULT 11
BB604671 627 bp mRNA linear EST 26-OCT-2000
LOCUS BB604671.1 NC_000000.1 Mus musculus cDNA clone IMAGE:3491991 5'
DEFINITION BB604671.1 GI:9156727
ACCESSION BB604671
VERSION BB604671
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 627)
NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@phs-research.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLN8537 row: a column: 16
High quality sequence stop: 530.
Location/Qualifiers
1. 627
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3491991"
/tissue_type="spontaneous tumor, metastatic to mammary
stem cell origin."
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-Sport6; Site: 1: SalI;
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 153 a 159 c 177 g 138 t
ORIGIN
Query Match 0.98; Score 26; DB 10; Length 627;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1942 GTCTGTGTCGAAGGCAAGTGCATCAA 1967
|||||
Db 95 GTCTGTGTCGAAGGCAAGTGCATCAA 120
|||||

RESULT 12
BB604671 247 bp mRNA linear EST 05-DEC-2000
LOCUS BB604671.1 NC_000000.1 Mus musculus cDNA clone E030005K13 5'
DEFINITION BB604671.1 GI:11556073
ACCESSION BB604671
VERSION BB604671
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 247)
NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@phs-research.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLN8537 row: a column: 16
High quality sequence stop: 530.
Location/Qualifiers
1. 247
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3491991"
/tissue_type="spontaneous tumor, metastatic to mammary
stem cell origin."
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-Sport6; Site: 1: SalI;
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 153 a 159 c 177 g 138 t
ORIGIN
Query Match 0.98; Score 26; DB 10; Length 627;
Best Local Similarity 100.0%; Pred. No. 0.47;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1942 GTCTGTGTCGAAGGCAAGTGCATCAA 1967
|||||
Db 95 GTCTGTGTCGAAGGCAAGTGCATCAA 120
|||||

RESULT 13
BB103177 849 bp mRNA linear EST 26-JUN-2001
LOCUS BB103177.1 NC_000000.1 Mus musculus cDNA clone IMAGE:5044493
DEFINITION BB103177.1 GI:11556073
ACCESSION BB103177
VERSION BB103177
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 849)
NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@phs-research.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLN8537 row: a column: 16
High quality sequence stop: 530.
Location/Qualifiers
1. 849
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5044493"
/tissue_type="spontaneous tumor, metastatic to mammary
stem cell origin."
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-Sport6; Site: 1: SalI;
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 153 a 159 c 177 g 138 t
ORIGIN
Query Match 0.98; Score 26; DB 10; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1700 TGGAGGCGCTGCCCCAGCTCAGCCTC 1724
|||||
Db 34 TGGAGGCGCTGCCCCAGCTCAGCCTC 58
|||||

Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C.,
Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shinagawa, A.,
Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tanaka
T., Toya, T., Watabiki, A., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Alizawa, K. et al. 2000)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1. 247
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="E030005K13"
/clone_lib="RIKEN full-length enriched, 0 day neonate
lung"
/tissue_type="lung"
/dev_stage="0 day neonate"
/lab_host="DH10B"
/notes="Site: 1: SalI; Site: 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGGAGAGAGCGCGCCGCACTCGAGTGTGTTTATTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGGAGAGATTCGAGTGTGTTTATTTTNN 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT 47 a 30 c 128 g 42 t
ORIGIN
Query Match 0.9%; Score 25; DB 10; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1700 TGGAGGCGCTGCCCCAGCTCAGCCTC 1724
|||||
Db 34 TGGAGGCGCTGCCCCAGCTCAGCCTC 58
|||||

RESULT 13
BB103177 849 bp mRNA linear EST 26-JUN-2001
LOCUS BB103177.1 NC_000000.1 Mus musculus cDNA clone IMAGE:5044493
DEFINITION BB103177.1 GI:11556073
ACCESSION BB103177
VERSION BB103177
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 849)
NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@phs-research.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLN8537 row: a column: 16
High quality sequence stop: 530.
Location/Qualifiers
1. 849
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5044493"
/tissue_type="spontaneous tumor, metastatic to mammary
stem cell origin."
/lab_host="DH10B"
/notes="Organ: lung; Vector: pCMV-Sport6; Site: 1: SalI;
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 153 a 159 c 177 g 138 t
ORIGIN
Query Match 0.9%; Score 25; DB 10; Length 247;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1700 TGGAGGCGCTGCCCCAGCTCAGCCTC 1724
|||||
Db 34 TGGAGGCGCTGCCCCAGCTCAGCCTC 58
|||||

```


ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BI103177
BI103177.1 GI:14554070
EST.
house mouse.
Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 849)
NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1122 row: a column: 06
High quality sequence stop: 743.

FEATURES
Source

Location/Qualifiers
1..849

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5044493"
/clone_lib="NCI CGAP Kid14"
/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library. |"
193 a 250 c 250 g 156 t

BASE COUNT
ORIGIN

Query Match 0.9%; Score 25; DB 13; Length 849;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1700 TGGAGCCCTGCCCGAGCTCAGCCTC 1724
|||||
Db 540 TGGAGCCCTGCCCGAGCTCAGCCTC 564

RESULT 14
BE666088

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

149119 MARC 4BOV Bos taurus cDNA 5', mRNA linear EST 25-APR-2001
464 bp mRNA linear EST 25-APR-2001
house mouse.
Mus musculus

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 464)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
Perteau, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keeler, J.W.

TITLE
JOURNAL
MEDLINE
COMMENT

Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390

Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.

PCR Primers
FORWARD: AGGAACAGCATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 62 row: B column: 2
Seq primer: ATTTAGGTGACACTATAG.

FEATURES
source

Location/Qualifiers
1..464

/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 4BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from day 20 and day 40
embryos." 73 a 157 c 163 g 71 t

BASE COUNT
ORIGIN

Query Match 0.8%; Score 24; DB 10; Length 464;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2812 AAGCCCCAGGAGCTGGACTTCTGC 2835
|||||
Db 215 AAGCCCCAGGAGCTGGACTTCTGC 238

RESULT 15
BI692986

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BI692986 664 bp mRNA linear EST 18-SEP-2001
603344521F1 NCI CGAP_Mam2 Mus musculus cDNA clone IMAGE:5372261 5',
mRNA sequence.
BI692986 GI:15655615
EST.
house mouse.
Mus musculus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 664)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11947 row: j column: 06
High quality sequence stop: 664.

FEATURES
source

Location/Qualifiers
1..664

/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5372261"
/clone_lib="NCI CGAP Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

BASE COUNT
ORIGIN

133 a 168 c 207 g 156 t

ORIGIN

Query Match 0.8%; Score 24; DB 13; Length 664;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2752 CTCAGTGTGTGGCCACGGAGGC 2775
|||||
Db 46 CTCAGTGTGTGGCCACGGAGGC 69

Search completed: May 2, 2003, 04:28:09
Job time : 3943 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 15.41 seconds
(without alignments)
1813.869 Million cell updates/sec

Title: US-10-009-332-1
Perfect score: 5164
Sequence: 1 MLLGLILTLAFAGTAGGFE.....DQCNLHRKPOELDFCVLRPC 950

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	2480.5	48.0	967	4	US-09-130-491-2
2	2113	40.9	905	4	US-09-369-364A-9
3	1917	37.1	837	4	US-09-122-126B-2
4	1900	36.8	608	4	US-09-130-491-13
5	1854	35.9	930	4	US-09-369-364A-2
6	1848.5	35.8	1882	4	US-09-369-364A-13
7	1847.5	35.8	930	4	US-09-122-126B-15
8	1798.5	34.8	874	4	US-09-369-364A-15
9	1607	31.1	551	4	US-09-130-491-16
10	1273	24.7	481	4	US-09-130-491-8
11	1264	24.5	518	4	US-09-369-364A-22
12	1248.5	24.2	1224	4	US-09-930-872-4
13	1217	23.6	1211	4	US-09-491-522-5
14	1202.5	23.3	997	4	US-09-369-364A-7
15	1152	22.3	1081	4	US-09-369-364A-17
16	1144	22.2	1205	4	US-09-491-522-11
17	1035	20.0	859	4	US-09-369-364A-5
18	662.5	12.8	245	4	US-09-369-364A-11
19	576.5	11.2	566	4	US-09-491-522-7
20	550.5	10.7	525	4	US-09-369-364A-21
21	441	8.5	491	4	US-09-930-872-2
22	380.5	7.4	450	4	US-09-369-364A-19
23	349	6.8	812	4	US-09-632-098-4
24	339	6.6	802	4	US-09-632-098-2
25	303	5.9	464	4	US-09-411-329C-14
26	301.5	5.8	616	4	US-09-608-790-1
27	301	5.8	462	4	US-09-411-329C-3

28	301	5.8	462	4	US-09-411-329C-17
29	301	5.8	529	2	US-08-836-442-3
30	285	5.5	621	4	US-09-026-001A-6
31	285	5.5	855	4	US-09-813-819-2
32	285	5.5	855	4	US-09-920-048-2
33	278	5.4	621	4	US-09-026-001A-18
34	277.5	5.4	814	4	US-09-813-819-4
35	277.5	5.4	814	4	US-09-920-048-4
36	276	5.3	769	1	US-08-243-542-4
37	276	5.3	769	1	US-08-477-407-4
38	276	5.3	769	1	US-08-484-355-4
39	274.5	5.3	751	2	US-08-836-443-3
40	273	5.3	613	4	US-09-026-001A-10
41	260	5.0	670	1	US-08-243-542-3
42	260	5.0	670	1	US-08-477-407-3
43	260	5.0	670	1	US-08-484-355-3
44	259	5.0	592	4	US-09-026-001A-14
45	248.5	4.8	1170	1	US-08-313-288B-20

ALIGNMENTS

RESULT 1
US-09-130-491-2
; Sequence 2, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-130-491-2

Query Match 48.0%; Score 2480.5; DB 4; Length 967;
Best Local Similarity 48.9%; Pred. No. 3.1e-198;
Matches 485; Conservative 154; Mismatches 251; Indels 101; Gaps 24;

QY	1	MLLGLILTLAFAGTAGGPEPEREVVPIRLDPDINGRYYWRGPDSDGQGLIFQITAF	60
DB	36	LLLLAALAVSALGRPEDEELVVP-ELE-----RAP---GHGTRULRHAF	81
QY	61	QEDFYLHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRCFYSGDVNAPDSF	114
DB	82	DQQLDELRLPDSFLAPGGTFLQNVGRKSGSEYPLP-----ETDLAHCFFSGTVNGDPSSA	136
QY	115	AAVSLCGGLRGAGFYGAEEVVISPLPNAS---APAAQRNSQGA-----HLLQ---RRGVPGG	165
DB	137	AALSCEGVGRGAYLLGEAYFIQPLPAASERLATAPEKPPAPLQFHLRLNRNQDVG	196
QY	166	PSG-----DPTSRC-----GVASG--WNPAIRALDPYKPRRAGFESRRRS	207
DB	197	TCGVVDEPRPTKATEDEDEGTEDEGPWS-----PQDPALGVGPQ-TGTGS	247
QY	208	GRAKREVSRPVYVETLVVADESVMKFGADLEHLLTLTLATARLYRHPISILNINIVV	267
DB	248	IRKREVSSRRYVETMLVADQSMAEFHGSLKHYLLTLFSVAARLYKHPISIRNSVLVV	307
QY	268	KVLLLRDSDGPKVTGNAALTFRNFCAMOKKLNKVDKHPYWDYTAILTRDOLCGATTC	327
DB	308	KILVIHDEQGPVTSNAALTFRNFCNQHNPPSDRDAEHYDVTAILTRDOLCGSQT	367

[illegible]

; TYPE: PRT

US-09-369-364A-9

Best Local Similarity 45.58; Pred. No. 1.3e-167;

[illegible]

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-122-1268-2

Query Match          37.1%; Score 1917; DB 4; Length 837;
Best Local Similarity 46.3%; Pred. No. 2.6e-151;
Matches 381; Conservative 123; Mismatches 267; Indels 52; Gaps 16;

QY 1 MLLGILTLAFAGTAGGFEPEERVVPIRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
DB 37 LLLLLLSPASRLASPLPREEIVPEKLINGSVL-----PGSGAPARLLCLRLOAF 88

QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLITGSGDLRRCFYSGDVNABPDSDFAAVSLC 120
DB 89 GETLLELEQDSGVQVEGLTWYLGQAPE-LLGSAEP-CYTLTGTINGDPESVASLHWD 145

QY 121 GG-LRGAFYGAEYVISPAPASAPAAQNSOGAHLQRRGVGPGSDPTSRGCVASG 179
DB 146 GGALLGVLOVAGALHQLPLEGGTPNSA--GGPCAHILRRK-----SPASQGGPMCNV--- 196

QY 180 WNPAILRALDPYKPRRAGFESRRSRGAKRFVSTPRVETVLYVADSWKVFHGAJLE 239
DB 197 -----KAPLGSPPRPR--RAKRFASLSRFEVETLVVADDKMAAFHGAJLK 239

QY 240 HYLTLTLATAARLYRHPISILNPINVVVKKLLLRDRDSGPKVTGNAALTILRNFCAMOKKL 299
DB 240 RYLLTVMAAAKAPKHPISIRNPVSLVRLVILGSGEGPQVGPSPAAQTLRSECAWQRL 299

QY 300 NKVSDKPIPEWDTRAILTRDQLCGATCDTLGMADVTGMDCKPRSCSVIEDDGLPSAFTT 359
DB 300 NTPEDSDPDHFDTAILTRDQLCGVSTCDTLGMADVTGTCDPARSCAIVEDDGLQSAFTA 359

QY 360 AHELGHVFNPHONVNVKCEWVFGLK-RANHHMSPTLIQIDRANPNWASACAAIITDFLDG 418
DB 360 AHELGHVFNMLHNSKPCISLNGPLSTRHVMAPVAHVDPPEPWPSCSARFITDFLDNG 419

QY 419 HGDCLLDQPSKPISLPDLPGASYTLISQOCELAFVGSKPCPYM-QYCTKLMCTGKAKQ 477
DB 420 YGCLLDKPEAPLHLVTPFPKDYDADQCQLTFGPDSTRHCPLPPLPPCAALWCSGHLNGH 479

QY 478 MVCOTREHPWADTSGEGKCLKLGACVERHNLKHVRD--GSWAKWDYPGPCSRTCGGG 535
DB 480 AMCTKPSHWADTTPCGPAQACMGGRCLHMDQLQDFNIPQAGGWGPWGPWDCSRTCGGG 539

QY 536 VLARRQCTNPTPANGKCYCEGVRYRYSNLENLPCPSSASGSKSFEEOCEAFNGYHNSIN 595
DB 540 VQFSRDCRTPVRNPGKCYCEGRTRFRSCNTEDCP--TGSALTFRREQCAA---YNNHRVD 595

QY 596 RLFL---AVAWPKYSGVSPRDKCLICRANGTGYFVILAPKVVDGTLGSPDSTSVCGQG 652
DB 596 LFKSFPQMDWMPRYTGVAPODQCKLTCQARALGYVYVLEPRVVDGTCPSPDSSSVCGQG 655

QY 653 KCJAKGCDNLGSKKRFKDCGCGGDNKSKCKVTGLFTPMHGYNFVVAIPAGASSIDIR 712
DB 656 KCHAGCDRLIGSKKKFKDCMVGCGDSCQSGSFRFRYGYNNVVTIPAGATHILVR 715

QY 713 QRGYKGLIGDNNYLLALKNSOGKYLNLGHFWSAVERDLVVKSL-LRYSGTGTAVESLQA 771
DB 716 QQGNPG--HRSIYLALKLPDGSYALNGEYTLMPSTDDVVLPGAVSLRYSGATAASETLSG 773

QY 772 SRPILPPLTVELVSVGKMTPPRVRYSFYFLKPEPRDKSKSHPKD 814
DB 774 HGPLAQPLTLQVLVAGNPQDTRLRYSFYFVRPTPTPTPTPD 816

RESULT 4
US-09-130-491-13
; Sequence 13, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:

```

```

; Sequence 2, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; - NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 930
; TYPE: PRT
; ORGANISM: mus musculus ADAMTS-5
; US-09-369-364A-2

Query Match      35.9%; Score 1854; DB 4; Length 930;
Best Local Similarity 43.1%; Pred. No. 5.4e-146;
Matches 369; Conservative 116; Mismatches 265; Indels 106; Gaps 13;

QY 90 GLTGSSDLRCFCYSGDVNAPDFAAVSLCGGLRGAGFYGAHYVISPPLNASAPAAQR 149
DB 121 GLSASSGHRGCHFCYGTGVDGSPRLAVFDLCGLDGGFAVVKHARYTLKPLRGSWAEYER 180
QY 150 -----NSOGALLORRG-----VPGPGSGDPTSRGCVASGWNPAIRALDPY 191
DB 181 IYDGSRRILHYNREGSFALPRASCETPASP----- 217
QY 192 KPRRAGFGESESRSSG-----RAKRFVSIPIRYVETLVAD 227
DB 218 -PQESPVSRSRRSALAPQLLDHSAFSPSGNAGPQTWRRRRRSISRARQVELLVAD 276
QY 228 ESMVKFHGADLEHLLTAAARLYRHPSTLNPINVVVKKVLLLRDSDPKVTGNAAL 287
DB 277 SSMARYGRGLOHLYLTWASIANRYSASHTENIRLAVKVVVLTDKDTSLEYSKNAAT 336
QY 288 TLNRCFAWKLNKYSDKHPEYWDTAILFTRODLGATCTDGLMADVTGMDPKRSCSV 347
DB 337 TLKNFCCKQHOHNGDDHEHYDAAILFTREDLCGHHSCTDGLMADVTGTCSPERSCAV 396
QY 348 IEDDGLPSAFTAHGLHVFNMHPDNVNVCEVFGKLRANHMSPFLIQRANPWSACS 407
DB 397 IEDDGLHAFTVAHEIGLLGLSHDDSKFCENFGTTEDKRLMSILTSIDASKPWSKCT 456
QY 408 AAIITDFLDGSHGDCLLDQPKPISLPDLPGASYTLSSQCELAFAFGVSKPCPYMOYCTK 467
DB 457 SATITEFLDDHGNCLLDLPRKQIILGPBELPGQTYDATQOCNLTFGPPEYSVCPGMDVCAR 516
QY 468 LMCTGKAGQWYQTRHPFADGTSCGEGKCLGKACVER-----HNLKRVDSGWAKW 522
DB 517 LNCVVROGMYCLTKKLPAVEGTPCGRGVCLQCKVDKTKKYISTSSH---GNMGSW 573
QY 523 DYPGCSRTCGGVOLARROCTNPTPANGKYCEGVRYKRSCLNEPCPSASGKSFRFE 582
DB 574 GPWGQCSRGCGGVQFAYHCHNPNAPRSGRYCTCKRAIYRSCSVTPCP---PNGK$FRHE 631
QY 583 QCEAFNGYHNSHTRITLAVAWYPKYSGVSPRDKCLICRANGTGYFYVYAPKVDVGLICS 642
DB 632 QCEAKNGYQSDAKGVKTFVWVYKAGYLPADVCKLTCAKGTGYVYVSPKVTDTGTECR 691
QY 643 PSTSVCVGCKIKAGCDNGLSKRFDKCGVCGGDNKSKKVTGLTKPMHGVNFVVAI 702
DB 692 PYSNSVVRGVRGTCGDDIGISGLQYDKCGVCGGDNSSCTKIITFNKSKGYTDVRI 751
QY 703 PAGASSIDIRQRYKGLIGDDNYLAKNSQGRYLLNGHVFVSAVERDLVWAGSLRYSGT 762
DB 752 PEGATHIKVYQKANDQTRFAYLALKKTKTEYLINGYIMISTSETIIDINGTVMNYSGW 811
QY 763 GTAVESL-----QASRPTELEPTVEVLSVGKMTPPRVKYSFYLKPEPREDSK---SHPKD 814
DB 812 SHRDDPLHGMGYSATKEI---LIVQILATDPTKALGVRYSEFVPKKTQKQNSVISHGNS 868

```

```

QY 815 PRGFSVLHNSVLSLNQVEQDDPPARVWAGSWGSCSASCSGLQKRAVDCRGSAQORT 874
DB 869 KVGPD---HSTQL-----QWVTGFWLACSRCTCDTGWHTRTVQCQ-DGNRKL 909
QY 875 VPACDAARHPVETOAC 890
DB 910 AKGCLLSQPSAFKOC 925

RESULT 6
US-09-369-364A-13
; Sequence 13, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1882
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-9
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (468)
; OTHER INFORMATION: Xaa = C
; NAME/KEY: MOD_RES
; LOCATION: (521)
; OTHER INFORMATION: Xaa = Y
; US-09-369-364A-13

Query Match      35.8%; Score 1848.5; DB 4; Length 1882;
Best Local Similarity 39.6%; Pred. No. 4.4e-145;
Matches 396; Conservative 144; Mismatches 340; Indels 121; Gaps 23;

QY 22 EREVVPPIRLDP-----DINGRRYVW-----RGPEDSGDQGLIFQITAFQ 61
DB 49 EYEIVSPIRVNALGEPPTNVHFKRTRRSINATDPWAFASSSSSTSSQAHYRLSAFG 108
QY 62 EDFYHLTPDQAFLAPAFSTEHLGVP-----LQGLTGSSDLRCFCYSGDVNAPDFAAV 117
DB 109 QQFLNLTANAGFIAPLFTVLLGTPGVNQTKEYSEEAELKHCFYK----- 155
QY 118 SLGGLRGAGFYGAHYVISPPLNASAPAAQRNSOGAHLIQRVGPSPGSDPTSRGVA 177
DB 156 RLQOYQLRANG-----RHQPLLRNEKHNHSDKKKTRARKWGERINLAGDV 202
QY 178 SGWNPAI-LRALDPYKPRRAGFGESESRSSRAKRFVSIPIRYVETLVVVADESQWVFHGA 236
DB 203 AALNSGLATEAFSAIGNKTDNTRKTRHRT---KRFLSYPREVEVLVADNRMVSYHGE 259
QY 237 DLEHYLLTLTAARLYRHPSTLNPINVVVKKVLLLRDSDGPKVTGNAALTILRNFCAWQ 296
DB 260 NLQHYLLTMSIVASITRDFSIGNLINIVIVNLIVHNEQDGSISFNAGTILKNFCOMQ 319
QY 297 KKLNVKSDKHPEYWDTAILFTRODLGCA--TTCDTLGMADVGTWCDPKRSCSVTDDGLPS 355
DB 320 HSNPGGCIHH---DTAVLLTRQDIDICRAHKCDTLGLAELGTICDPRYRSCSISDSGLST 375
QY 356 APTAHGLGHVFNPHDNVNVCEVFGKLRANHMSPFLIQRANPWSACSAAITDFL 415
DB 376 APTAHGLGHVFNPHDNNDNNKKEE-GVKSPQHVMAEPLNFYINPMWMSKSKRYITEFL 434
QY 416 DSGHGDCLLDOP-SKPISLPEDLPGASYTLSSQCELAFAFGVSGKPCPYMOYCTKLWC---T 471
DB 435 DTGEGCELLNEPESRPYLPVLQPLGILYNNKQELIFGPGSQVCPYMMQCRRLWCNNVN 494

```

Db	247	-AGGSGPQWRRRRS-----ISRARQVELLLVADASMARLYXGRGLQHVLTLASIAN	299
Qy	251	RLYRHPSLNPINTVWVVKVLLLRDRDSGPKVTGNAALTRLNFCAWKAKLVKSDKHPEW	310
Db	300	RLYSHASINERLAVKVVVVLVDGDKLSLEYSKNAATLLKNFCKWQHQLQGLGDDEHY	359
Qy	311	DYAILFTRODLGCGATCDTLGMADYGTMCDDPKRSCSVTEDDGLPSAFTTAHELGHVFNMP	370
Db	360	DAAILFTREDLCGHSHSCDTLGMADVGTICSPERSCAVIEDDGLHAAFTVAHEIGLLGLS	419
Qy	371	HDNVKCEVFGKLRANHMSPTLIQIDRANPWSACSAALITDFELDSHGDCILLDQSPK	430
Db	420	HDSKFCBEETFGSTEDKRLMSSILTSIDASKPWSKCTSATITFELDDGHGNCILLDLPKRO	479
Qy	431	ISLPEDLPASVYTLSSQOCELAFLGVSCKPCPYQYCTKLWCTGKAKGVQMVQTRHFFWADG	490
Db	480	ILPEELPQQTVDATQOCLNTFFGPEYSVCPGMDVCARLWCVAVRGQGMVCLTKKLPAAVEG	539
Qy	491	TSCGEGKLCGLKACVER-----HNLAHKRVDCGSWAKWDPYGPCSRTCGGVQVLARQCNT	545
Db	540	TPCGKGRICLQKCKVDTKTKKYYSSTSH---GNWSGWSGWCQSKSGCGGVQVYARHCNN	596
Qy	546	PTPANGKYCEGRVRYKRSNCNLBPCPSSASGKSFRBEOCEAFNGYHNHSTNRLTLAVAWVP	605
Db	597	PAPRNNGRYCTKRALYRSCSLMPCP---PNGKSFRHEOCEAKNGYQSDAKGVKTFVEWVP	654
Qy	606	KYSGVSPRDKCLICRANGTGYFYVLAPKVVDTGLTCSDPDSTVCVQGRKCKAGDCGNLGS	665
Db	655	KYAGVLPADVCLTKCRAGTGYVYVSPKVTDTGECRYSNSVCYRGCYRVCSDGLIGS	714
Qy	666	KKRFDKCGVCGGDNKSCKVYTGFLTPRMHGYNFVAIPAAGASSIDIIRQRYKGLIGDDNY	725
Db	715	KLQYDKCGVCGGDNSSCTKIVGTFNKSKGYTDVRIPEGATHIKVROFAKQDQTRPTAY	774
Qy	726	LALKNSQGYLLNGHFVYASVERDLVYKGSLLRYSTGTVAESL-----QASRPILLEPLT	780
Db	775	LALKKNGEYLLNGKYMISTSETIIDINGTVMNYSGWSHRDDFLHMGYSATKEI---LI	831
Qy	781	VEVLSVGKMTPPREYSFYLPKPEPRDKSHKPDPRGPSVLJNSVLSL-SNOYEQPDPR	839
Db	832	VQLIATDPTKLDVIRYSFVFPKK-----STPK-----VNSTSHGSKNGVSHTSQP	877
Qy	840	PARWAGSWGPCSCSGSLGQKRAVDCRGSAGQRTVPACDAAHRPVETQAC	890
Db	878	--OWTGPWLACSRCTDGTGHTRTVQCO--DGNRKLAKGCPLSQRPSAPKOC	925

[illegible][illegible]

; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-130-491-16

Query Match 31.1%; Score 1607; DB 4; Length 551;
Best Local Similarity 48.5%; Pred. No. 1e-125;
Matches 282; Conservative 108; Mismatches 152; Indels 40; Gaps 10;

QY 377 CREVFGLKRLANRMSPTLIQIDRANPWSACSAAITDFLDGSHGDCLLDOP-SKPISLPELPCASY 436
DB 1 CASLNGVSDSHLMSLSLSDHSPWSPCSAYMVTSLFDNGHGECLMDKPNPIKLPSP 60
QY 437 LFGASTYLSQOCELAFGVGSKPCP-YMQYCTKLWCTGKAKGQMYCOTRHFPHWADTSCGE 495
DB 61 LPTGLYDANRQOFTFEESKHCPCDAASTCTTLWCTGTSGGLLVCTQTRHFPWADTSCGE 120
QY 496 GKLCAGACVERHNKHKH--RVDSWAKWDPYPCSRCTCGGVQVQLARRQCTNPTPANG 552
DB 121 GKWCVGSKCVNKTDK-KHEATPVHSGSWGPGWGDSCRTCGGVQYTMRECONPVKNGG 179
QY 553 KYCEGVRYKRYSCNLEPCPSSASGSKSFREEQCEAFNGYHNSTNRLTLAVAWPKYSGVSP 612
DB 180 KYCEGRVRYSCNIEDPCDN-NGKTFREEQCEAHNEFSKASFGNEPTVEMTPKYAGVSP 238
QY 613 RDKCKLICRANGTYFVVLAPKVVDCITLSPDSTSVCVQKCIKAGCDNGLSKRRFDC 672
DB 239 KDRCKITCEAKGIGYFVLQPKVVDDTGLSPDSTSVCVQGVQCVKAGCDRIIDSKKFKDC 298
QY 673 GVCGGNCKSKVTGLFTKPMHGYNEFVAIPAGASSIDIRQYKGLIGDDNYLAKNSQ 732
DB 299 GVCGGNGSTCKMKGIVTSTPCYHDIPTIPAGATNIEVKHNRQSRNNGSFLATRAAD 358
QY 733 GYLLNGHFFVSAVERDLVVGSLRYSGTGTAVESLQASRPILPTVEVLVSGKMTTP 792
DB 359 GTYILNGNFTSLTEQDLYTKGTVLRYSGSAALERISFSLKEPLTTOVLVGHALRP 418
QY 793 RVRYSFYLPKEPREDKSSHPKDPGRPSVLNSVLSI-SNQVEQPDPRPPARWAGVSGPCS 852
DB 419 KIKFTYPMKKTES-----FNAIPTS-----EWVTEEGECS 451
QY 853 ASCSGSLQKRAVDCRSAGORTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCKSKC 909
DB 452 KTCGSGWQRVWQCRDINGH---PASECAKEVKPASTRCPADLPCHPMQVGDWSPCKSKC 508
QY 910 GRGFORSLKCVGHGGRLLARDQCNLHRKPQE-LDFCVLRPC 950
DB 509 GKGYKKRILKCVSHDGVLSNESCDPLKKPKHYIDFCTLQOC 550

RESULT 10
US-09-130-491-8
; Sequence 8, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 481

[illegible]

RESULT 12
US-09-930-872-4
; Sequence 4, Application US/09930872
; Patent No. 648388
; GENERAL INFORMATION:
; APPLICATION NO.: 09-930-872-4
; PRIORITY NO.:
; FILING DATE: 09/09/2009
; INVENTOR(S):
; APPLICANT:

```

; TITLE OF INVENTION: No. 5448388e1 Human Proteases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0219-USA
; CURRENT APPLICATION NUMBER: US/09/930,872
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,852
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1224
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-930-872-4

Query Match          24.2%; Score 1248.5; DB 4; Length 1224;
Best Local Similarity 32.9%; Pred. No. 2.7e-95;
Matches 33%; Conservative 12%; Mismatches 386; Indels 165; Gaps 38;

```

Db	148	QGSLSHRHNSVALSTCGLSGMI	TEADYFLRLPLSHLSWKLGR	AQGSFSSHVLYR	207
QY	161	G----	VPGFSGDPTSR-----	CGVAGSNPA-----	ILRA 187
Db	208	STEPHAPGASVLSTRWELAHQ	PLHSSDLRLGLPKQKHFCGR	KRYMPQPPKEDFLIL	267

QY 188 LDPKRRAGGSRSSRRSRRAKRFVSIPIRYVETLVVADESVMKPHG-ADLEHVLTL 246
Db 268 PDEYK---SCLRHKRSLLSHRNEEL-----NVTLLVVDDKMMQNHENITTVLIL 319
QY 247 ATARLYRHPSILNPINVVVKKVLLLRDRSDGPKVTGNAALIRNFCAWKKL-NKVSOK 305
Db 320 NMVSALFKDTCIGNINIAVLGILLDEBQGLVISHADHTLSSPCQWQSLMGKRGDTR 379
QY 306 HPEYWDTAILFTRODLGG--ATTCDILGMADVCTMCDPKRSCVIEDDGLPSAFTTAHEL 363
Db 380 H----DHAILLTGIDICSWKNEPCDITGLFAPISGMC SKYRSCCTINEDTGLGLAFTAHES 435
QY 364 GHVFNPHDNV-KVCEVFEKLRANHMSPTLIQIDRANPWSACSAAIITDFLDSHGDC 422
Db 436 GHNFMTIHDEGNCKKSEG-----NIMSTLAGRNGVSWSPCSQRYLHKFLSTQAIC 490
QY 423 LLDQPSKPI---SLPEDLPAGSYTLQQCELAFAFGVSKPCPY---MOYCTKLMC---TGKA 474
Db 491 LADQP-KPVKEYKYPEKLPGELYDANTQKWQFGEKAKLCMLDFKDKICKALWCHIRGRK 549
QY 475 KGWVCOTRIHPHADTSCGEGKLCILKGACVERHNLNKHVRDGSMAKWDPYGPCSRTCGG 534
Db 550 -----CETKMPAABGTICGHDWCRGQCVKYGDGPKPTHGHSWSSWSPCSRTCGG 604
QY 535 GVQLARQCTNPTPANGKYGCEGVVRYRSCNLEPCSPSSAGSKSFREEQC-----EAFNG 589
Db 605 GVSHRSLCTNPKRSHGKCEGSTRTLKLCNSQKCPRDS--VDFRAAQCAEHNSRRFRG 662
QY 590 YNHNTRLTALAVAWPKYSGVSPRDKCKLICRANGTGYFYVLAPKVVDGTLCSPDSTSV 649
Db 663 RHYK-----WKP-YTQVEDQDLKLYCIAEGDFEFLSLNKKVKGDTGTPCSEDSRNV 712
QY 650 VOGKICRACGDNGLSKRDKCVCVCGDNKSKCKVTGLFTKPMH--GYNFVVAIPAGAS 707
Db 713 IDGICERVCGDNVJGSDAIVEDVCGVCGNNSACTIHRGLTKHHHTNQYHMTIPSGAR 772
QY 708 SIDIRQYKGLIGDNDYALNLSQGYLLNGHEFVVSVAVERDLVVKGSLLRYSGTGTAVE 767
Db 773 SIRIVEMNV-----STSVISVRNALRYYLNGHTVDMWPGR-YKFSGTTFDYRRSYNEPE 826
QY 768 SIQASRPLEPLTVELSVLWGTTPRYRYSFYLKPEPREDKSSHPKDPGRPS-----819
Db 827 NLIATGPTNETLIVELLFQGR--NPGVAWEYSMPR-----LGTEKQPPAQPSYTAIVRS 879
QY 820 -----VLHNSVLSLSNOVE-----QPDDR-----PARKVAGSNGPC 851
Db 880 ECSVSCGGQMTVRGCGYRDLKFQVNSFCNPKTRPVTLGVPCVKSACPPSWSVGNNSAC 939
QY 852 SASCGSGLOKRAVDC--RGSAGQRTVPA--CDAARHPVETOAC-GEPC-PTWELSAWSPC 905
Db 940 SRTCGGAQSRPVQCTRRVHYDSEVPVPSLCPQP-APSSRQACNSQSCPPAWSAGPWAEC 998
QY 906 SKSCGRGQRRSLKCVGHG---GRLARDOCNLHRKPOELDFCVLRPC 950
Db 999 SHTCGKGRKRAVACKSTNPSARAQLLPDAVCTSEPKPRMHEACLLQRC 1047

RESULT 13
US-09-491-522-5
; Sequence 5, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Colige, Alain
; APPLICANT: Lapierre, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE.
; NUMBER OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York

STATE: NY
COUNTRY: USA
ZIP: 10036-2811
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/491.522
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/886.333
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Abrams, Samuel B
REGISTRATION NUMBER: 30,605
REFERENCE/DOCKET NUMBER: 8389-0060-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1211 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-491-522-5
Query Match 23.6%; Score 1217; DB 4; Length 1211;
Best Local Similarity 33.1%; Pred. No. 1.1e-92;
Matches 347; Conservative 133; Mismatches 405; Indels 162; Gaps 47;
QY 9 LAFAGRTAG---GFEPERVVVPIRLDPD-----INGR-----YYW 42
Db 39 LAAADPPGGLPGHGAERILAVPVRTDAQRLVSHVSAATSRAGVARRARVPTSPFP 98
QY 43 RGEDSGDGLFIQTAQEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRRCF 102
Db 99 GNEEEFGSHLFYVTVFGDRLHLRLRPARLVAPGATMEWGE--KGITRVEPLLGSL 156
QY 103 YSGDVN--AEPSFAAVSLCGGLRGAFYGAIEYISPLPNASAPAAQNSQG-AHLQOR 159
Db 157 YGVDVAGLAFASS-VALSNCGLAGLIRMEEEFFTEPLEKGL--AAQEAEGRVHVYR 213
QY 160 RGVPGGPGSDPTSRGVASGWNPAIRLALDPYKPRAGFGESRRSRRAKRFVSIPIRY 219
Db 214 RPPTSPPLGGPQA---LDTG---ASLDSLSLS--RALGVLEEHSRRRRRHAADDDY 266
QY 220 VE-TLVVADESVMKPHGAD-LEHVLTLTATAARLYRHPSILNPINVVVKKVLLLRDRS 277
Db 267 NEIVLGVDDSVVQFHGKEHVQKYLITLNMVNEIYHDESGLAHINVLVRIILSYGKS 326
QY 278 GPKV-TGNAALTIRNFCAMQKLNKYSKDHPEYWDTAILFTRODLGATCTDTFLGMADV 336
Db 327 MSLEIGNPSQSLNVCRWAYLQPKPTDGHDEYDHAIFLTRQDF-GPSQM-Q--GYAPVT 383
QY 337 TMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNV-KVCEVFEKLRANHMSPTLI 395
Db 384 GMCHPVRCTLNHEDGFSFSAFVVAHETGHVLMGHEDGQGNRCGD---EVLGLSMAPLVQ 440
QY 396 QIDRANPWSACSAAIITDFLDSHGDCCLDQP---SKPISLPEDLPASVYTLSSQCELA 452
Db 441 AAFERFHWRSQOELSRYLHS--YDCLLDODPAHDWP-ALPQ-LFGLHYSMNEQCRFDF 496
QY 453 GVSGKPCPYQY---CTKLMCTGKAKGMVQCOTRHFHPWADGTSCGEGKLCILKACV----505
Db 497 GLGYMCTAFRTFDPCKQLWCS-HPDNPYFCKTKGKPPDCTMCAPGKHCFKGHCILWLP 555
QY 506 ---ERHNLNKHVRDGSMAKWDPYGPCSRTCGGQVQLARRQCTNPTPANGKYGCEGVV 562

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:13:53 ; Search time 49 Seconds
(without alignments)
2583.430 Million cell updates/sec

Title: US-10-009-332-1
Perfect score: 950
Sequence: 1 MLLGLTLAPAGTAGGFE.....DQCNLHRKPOELDFCVLRPC 950

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	950	100.0	950	22	AA62299 Human metalloprotease #2.
2	931	98.0	950	23	AAE22541 Human metalloprotease #2.
3	673	70.8	952	23	AAU74751 Human metalloprotease #2.
4	635	66.8	928	23	AAU72899 Human metalloprotease #2.
5	345	36.3	367	23	AAE22542 Human metalloprotease #2.
6	300	31.6	321	23	AAE22540 Human metalloprotease #1.
7	177	18.6	381	21	AAE21261 Human metalloprotease #1.
8	154	16.2	505	21	AAE21257 Human metalloprotease #1.
9	17	1.8	727	20	AAW78435 Human ADAMTS-1 pro
10	17	1.8	896	21	AAE21265 Mouse metalloprotease

11	17	1.8	950	20	AA49501 Human METH1 proteol
12	17	1.8	950	21	AA53899 Amino acid sequenc
13	17	1.8	950	22	AA73549 Human ADAM-type me
14	17	1.8	950	22	AA50002 Human METH1 - Homo
15	17	1.8	967	19	AAW80285 Human integrin lig
16	17	1.8	967	20	AAW04142 Human tango-71 pro
17	17	1.8	967	20	AAW78189 Human secreted pro
18	17	1.8	968	22	AA50011 Protein; SEQ ID 12
19	14	1.5	180	22	AA95532 Human protein sequ
20	13	1.4	107	22	AA004896 Human polypeptide
21	13	1.4	625	23	AAW48394 Rat aggrecanase
22	13	1.4	837	20	AAW75425 Human aggrecan deg
23	13	1.4	837	21	AAW9429 Human PRO1563 (UNQ
24	13	1.4	837	22	AA78228 Human aggrecanase-
25	13	1.4	837	22	AAU29199 Human PRO polypept
26	13	1.4	837	22	AA66178 Human metalloprote
27	13	1.4	840	21	AAE21256 Human metalloprote
28	13	1.4	874	22	AAE2287 Murine ADAMTS-9 am
29	13	1.4	947	22	AAE86950 Human metalloprote
30	13	1.4	958	21	AAE21255 Human metalloprote
31	13	1.4	1073	21	AAE21264 Human metalloprote
32	13	1.4	1602	23	ABG30702 Human aggrecanase
33	13	1.4	1629	23	ABG30703 Human aggrecanase
34	13	1.4	1629	23	AAO14448 Human ADAMTS-SI pr
35	13	1.4	1882	22	AAE72286 Human ADAMTS-9 am
36	13	1.4	1907	23	AAU77133 Human protease #12
37	13	1.4	1916	23	AAE19173 Human metalloprote
38	13	1.4	1934	22	AAE72286 Human metalloprote
39	13	1.4	1935	23	AAU72896 Human metalloprote
40	12	1.3	12	21	AAE21273 Human ADAMTS 1 Zn
41	12	1.3	12	21	AAE21277 Human ADAMTS 9 Zn
42	12	1.3	98	21	AAE21262 Rat metalloprote
43	12	1.3	481	20	AAU04145 Rat tango-76 prote
44	12	1.3	905	22	AAE72284 Murine ADAMTS-8 am
45	11	1.2	58	22	AAE90625 Human secreted pro

ALIGNMENTS

RESULT 1
AA62299
ID AAG62299 standard; protein; 950 AA.

XX AAG62299;

XX AC

XX 23-AUG-2001 (first entry)

DE Human metalloprotease MDT56 protein.

XX Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;

XX osteopathic; antiarthritic.

OS Homo sapiens.

XX WO200134785-A1.

XX 17-MAY-2001.

XX 10-NOV-2000; 2000WO-JP07917.

XX 11-NOV-1999; 99JP-0321740.

PR 16-MAY-2000; 2000JP-0144020.

XX (YAMA) YAMANOUCHI PHARM CO LTD.

XX (KAZU-) KAZUSA DNA RES INST.

PI Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;

XX WPI; 2001-343602/36.

DR N-PSDB; AA41003.

XX Metalloprotease with aggrecanase activity for treating joint diseases

PT especially osteoarthritis -
XX Claim 1; Page 56-60; 85pp; Japanese.
PS
XX This invention relates to a metalloprotease with aggreganase activity.
CC The invention includes protein and DNA sequences of the metalloprotease,
CC vectors containing the DNA, host cells transformed by the vectors, and
CC antibodies directed against the metalloprotease. The antibodies, protein
CC and DNA sequences can be used in the treatment and prevention of joint
CC diseases, particularly osteoarthritis. The treatment may result in
CC osteopathic and antiarthritic activity. The present sequence represents
CC the metalloprotease of the invention termed MDT56.
XX
SQ Sequence 950 AA;

Query Match 100.0%; Score 950; DB 22; Length 950;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 950; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPDSDGGLIFQITAF 60
Dy 1 MLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPDSDGGLIFQITAF 60
Qy 61 QEDFYHLTPDAQFAPAFSTHGLVPLQGLTGSSDLRRCFCYSGDVNAEPDSFAAVSLC 120
Dy 61 QEDFYHLTPDAQFAPAFSTHGLVPLQGLTGSSDLRRCFCYSGDVNAEPDSFAAVSLC 120
Qy 121 GGLRGAFYGAERVIISPLNASAPAAQRNSOGAHLQRRGVPGPSGDPSTRCGVASGW 180
Dy 121 GGLRGAFYGAERVIISPLNASAPAAQRNSOGAHLQRRGVPGPSGDPSTRCGVASGW 180
Qy 181 NPAILRALDPKPRRAGFSGRSRRGRKRFVSIPIRYVETLVVADESVMKFGADLEH 240
Dy 181 NPAILRALDPKPRRAGFSGRSRRGRKRFVSIPIRYVETLVVADESVMKFGADLEH 240
Qy 241 YLLTLLATAARLYRHPSTLNPINIVVVKVLLLRDRDSGPKVTGNAALTLLNFCAWOKKLN 300
Dy 241 YLLTLLATAARLYRHPSTLNPINIVVVKVLLLRDRDSGPKVTGNAALTLLNFCAWOKKLN 300
Qy 301 KVSQKHPEYWDATILFTRQDLCGATTCDTLGMADVGMCDPKRSCSVIEDGLPSAFTTA 360
Dy 301 KVSQKHPEYWDATILFTRQDLCGATTCDTLGMADVGMCDPKRSCSVIEDGLPSAFTTA 360
Qy 361 HELGHVFNHNDNVKVEEVEFGKLRANHMSPTLIQIDRANPWSACSAAIITDFLDSGHG 420
Dy 361 HELGHVFNHNDNVKVEEVEFGKLRANHMSPTLIQIDRANPWSACSAAIITDFLDSGHG 420
Qy 421 DCLDQPSKPIISLPEDLPASVYTLISQCELAFGVGSKPCPYMQYCTKLWCTGKAKGQWVC 480
Dy 421 DCLDQPSKPIISLPEDLPASVYTLISQCELAFGVGSKPCPYMQYCTKLWCTGKAKGQWVC 480
Qy 481 QTRHPPWADGTSCGEGKLCILKAGACVERHNLNKRVDSWAKWDYGPSCRTCGGVOLAR 540
Dy 481 QTRHPPWADGTSCGEGKLCILKAGACVERHNLNKRVDSWAKWDYGPSCRTCGGVOLAR 540
Qy 541 RQCTNPTPANGKYCEGVYRVYRSCNLEPCPSSASGKSFREEOCEAFNGYNHSTNRLTLA 600
Dy 541 RQCTNPTPANGKYCEGVYRVYRSCNLEPCPSSASGKSFREEOCEAFNGYNHSTNRLTLA 600
Qy 601 VAWVPKYSGVSPRCKKILCRANGTGYFYVLAPKVDGTLCSPDSTSVCGQKCIKAGCD 660
Dy 601 VAWVPKYSGVSPRCKKILCRANGTGYFYVLAPKVDGTLCSPDSTSVCGQKCIKAGCD 660
Qy 661 GNLGSKKRFDCGCGGDKNCKKVTGLFTKPMHGYNFVVAIPAGASSIDIRQRYKGLI 720
Dy 661 GNLGSKKRFDCGCGGDKNCKKVTGLFTKPMHGYNFVVAIPAGASSIDIRQRYKGLI 720
Qy 721 GDDNYLALKNSQGYLLNGHFVWSAVERDLVVKGSLRLRYSGTGTAVESLQASRPILPLT 780
Dy 721 GDDNYLALKNSQGYLLNGHFVWSAVERDLVVKGSLRLRYSGTGTAVESLQASRPILPLT 780
Qy 781 VEVLVSGKMTTPRRYRYSYLPKEPREDKSSHPKDPGRGSPVLSLNSNOVEQDDRRPP 840
Dy 781 VEVLVSGKMTTPRRYRYSYLPKEPREDKSSHPKDPGRGSPVLSLNSNOVEQDDRRPP 840

Db 781 VEVLVSGKMTTPRRYRYSYLPKEPREDKSSHPKDPGRGSPVLSLNSNOVEQDDRRPP 840
Qy 841 ARWAGSWGSPCSASCGSLQKRAVDCRGSAQRTVPACDAARHPVETQACGECPTWELS 900
Dy 841 ARWAGSWGSPCSASCGSLQKRAVDCRGSAQRTVPACDAARHPVETQACGECPTWELS 900
Qy 901 AWSPCSKSGRGRFQRSLKCVGHGRLRLARQCNLHRKPQELDFCVLRPC 950
Dy 901 AWSPCSKSGRGRFQRSLKCVGHGRLRLARQCNLHRKPQELDFCVLRPC 950

RESULT 2
AAE22541
ID AAE22541 standard; Protein; 950 AA.
XX
AC AAE22541;
XX
DT 26-JUL-2002 (first entry)
XX
DE Human protease #2.
XX
KW Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme.
XX
OS Homo sapiens.
XX
PN WO200226949-A2.
XX
PD 04-APR-2002.
XX
PF 27-SEP-2001; 2001WO-US30350.
XX
PR 29-SEP-2000; 2000US-236689P.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Friddle CJ, Hilbun E;
DR WPI; 2002-372123/40.
DR N-PSDB; AAD35569.
XX
PT Novel nucleic acid encoding a human protease, useful as a hybridization
PS probe for screening libraries and assessing gene expression patterns -
PS Claim 6; Page 36-38; 41pp; English.
XX
CC The present sequence is novel human protein (NHP), human protease.
CC NHPs share structural similarity with animal proteases particularly
CC zinc metalloproteases. Sequences of the invention are useful in
CC therapeutic, diagnostic and pharmacogenomic applications. NHP
CC polynucleotides are used as hybridisation probes for screening
CC libraries and assessing gene expression patterns. They can also be
CC used for treating related biological disorders such as obesity, high
CC blood pressure, arthritis, connective tissue disorders and infertility.
CC They are also used in gene therapy.
XX
SQ Sequence 950 AA;

Query Match 98.0%; Score 931; DB 23; Length 950;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 931; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 EPEREVVVPIRLDPDINGRRYYWRGPDSDGGLIFQITAFQEDFYHLTPDAQFAPAF 79
Dy 20 EPEREVVVPIRLDPDINGRRYYWRGPDSDGGLIFQITAFQEDFYHLTPDAQFAPAF 79
Qy 80 STEHLGVPLQGLTGSSDLRRCFCYSGDVNAEPDSFAAVSLCGLRGAFYGAERYVISPL 139
Dy 80 STEHLGVPLQGLTGSSDLRRCFCYSGDVNAEPDSFAAVSLCGLRGAFYGAERYVISPL 139
Qy 140 PNASAPAAQRNSOGAHLQRRGVPGPSGDPSTRCGVASGNPAILRALDPKPRRAGFG 199
Dy 140 PNASAPAAQRNSOGAHLQRRGVPGPSGDPSTRCGVASGNPAILRALDPKPRRAGFG 199

```
Db 140 PNASAPAAQRNSOGAHLQRRGVPGPGSDPTSRGCVASGWNPAILRALDPYKPRRAGFG 199
QY 200 ESRRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPFSIL 259
Db 200 ESRRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPFSIL 259
QY 260 NPINIVVVKVLLLRDRSDSGPKVTGNAALTLRNFCAQWKLNKVDKHPYWDTAILFTRQ 319
Db 260 NPINIVVVKVLLLRDRSDSGPKVTGNAALTLRNFCAQWKLNKVDKHPYWDTAILFTRQ 319
QY 320 DLGCAATCTDLGADVTGMDCKRSCSVIEDDGLPSAETTAHELGHVFMHPDNVKNVCEE 379
Db 320 DLGCAATCTDLGADVTGMDCKRSCSVIEDDGLPSAETTAHELGHVFMHPDNVKNVCEE 379
QY 380 VFGKLRANHMSPPTLIQIDRANPWSACSAAITDFLDSHGDCCLLDQPSKPSISLPEDLP 439
Db 380 VFGKLRANHMSPPTLIQIDRANPWSACSAAITDFLDSHGDCCLLDQPSKPSISLPEDLP 439
QY 440 ASYTLQOCELARFGVSKPCPYMOYCTKLWCTCKAKGQWVCOTRFPWADGTSCEGKLC 499
Db 440 ASYTLQOCELARFGVSKPCPYMOYCTKLWCTCKAKGQWVCOTRFPWADGTSCEGKLC 499
QY 500 LKACVERHNLKRRVDSWAKWDPYPCGSRCTCGGGVQLARRQCTNPTPANGKYCEGVR 559
Db 500 LKACVERHNLKRRVDSWAKWDPYPCGSRCTCGGGVQLARRQCTNPTPANGKYCEGVR 559
QY 560 KYRSCNLEPCPSASGKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLI 619
Db 560 KYRSCNLEPCPSASGKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLI 619
QY 620 CRANGTGYFVVLAPKVVVDGTLQSPDSTSVCGQKCIKAGCDGNLGSRRFCKGVCGGDN 679
Db 620 CRANGTGYFVVLAPKVVVDGTLQSPDSTSVCGQKCIKAGCDGNLGSRRFCKGVCGGDN 679
QY 680 KSKCKVTGLTKPMHGYNEFWALPAGASSIDIRQGYKGLIGDDNYLALKNSQGYLLNG 739
Db 680 KSKCKVTGLTKPMHGYNEFWALPAGASSIDIRQGYKGLIGDDNYLALKNSQGYLLNG 739
QY 740 HFVYSAVERDLVVKSLRYSGTGAVESLQASRPILPTVEVLSVKMTPPRVRYSFY 799
Db 740 HFVYSAVERDLVVKSLRYSGTGAVESLQASRPILPTVEVLSVKMTPPRVRYSFY 799
QY 800 LPKEPRDKSSHPRGSRVLSLNSVLSNOVEOPDPPARWAGSWGSCSASGSGL 859
Db 800 LPKEPRDKSSHPRGSRVLSLNSVLSNOVEOPDPPARWAGSWGSCSASGSGL 859
QY 860 QKRAVDCRGSAQRTVPACDAARHPVETQACGEPCTWELSAWSPCSKSGRGFORSLK 919
Db 860 QKRAVDCRGSAQRTVPACDAARHPVETQACGEPCTWELSAWSPCSKSGRGFORSLK 919
QY 920 CVHGGELLARDQCNLHRKPQELDFCVLRPC 950
Db 920 CVHGGELLARDQCNLHRKPQELDFCVLRPC 950

RESULT 3
AAU74751
ID: AAU74751 standard; Protein; 952 AA.
XX
AC AAU74751;
XX
DT 09-APR-2002 (first entry)
XX
DE Human protease PRTS-11 protein sequence.
XX
KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis.
XX
OS Homo sapiens.
```

```
XX WO200198468-A2.
PN
XX 27-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-US191178.
XX
PR 16-JUN-2000; 2000US-2123336P.
PR 22-JUN-2000; 2000US-213955P.
PR 29-JUN-2000; 2000US-215396P.
PR 07-JUL-2000; 2000US-21821P.
PR 14-JUL-2000; 2000US-218946P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
XX Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Walla NK, Yao MG, Lu DM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Rallick DA;
XX
DR WPI; 2002-090437/12.
DR N-PSDB; ABK12894.
XX
XX Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT proliferative (e.g. cancer) disorders -
XX
PS Claim 1; Page 144-146; 177pp; English.
XX
XX The present invention relates to twenty one new human proteases,
CC referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC myocardial infarction, autoimmune/inflammatory e.g. acquired
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC endometriosis disorders. Numerous other examples of each disorder are
CC given in the specification. The present protein sequence represents
CC the human protease PRTS-11 protein of the invention.
XX
SQ Sequence 952 AA;
XX
Query Match 70.8%; Score 673; DB 23; Length 952;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 793; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 20 EPREVVVPIRLDPDINGRRYWRGPGSDGGLIFQITAFQEDFYHLTTPDAFLAPAF 79
Db 20 EPREVVVPIRLDPDINGRRYWRGPGSDGGLIFQITAFQEDFYHLTTPDAFLAPAF 79
QY 80 STEHGLVPLQGLTGSSDLRRCFYSGDVNAEPDFAAAYSLGGLRGAFYGAEVISPL 139
Db 80 STEHGLVPLQGLTGSSDLRRCFYSGDVNAEPDFAAAYSLGGLRGAFYGAEVISPL 139
QY 140 PNASAPAAQRNSOGAHLQRRGVPGPGSDPTSRGCVASGWNPAILRALDPYKPRRAGFG 199
Db 140 PNASAPAAQRNSOGAHLQRRGVPGPGSDPTSRGCVASGWNPAILRALDPYKPRRAGFG 199
QY 200 ESRRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPFSIL 259
Db 200 ESRRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPFSIL 259
QY 260 NPINIVVVKVLLLRDRSDSGPKVTGNAALTLRNFCAQWKLNKVDKHPYWDTAILFTRQ 319
Db 260 NPINIVVVKVLLLRDRSDSGPKVTGNAALTLRNFCAQWKLNKVDKHPYWDTAILFTRQ 319
QY 320 DLGCAATCTDLGADVTGMDCKRSCSVIEDDGLPSAETTAHELGHVFMHPDNVKNVCEE 379
Db 320 DLGCAATCTDLGADVTGMDCKRSCSVIEDDGLPSAETTAHELGHVFMHPDNVKNVCEE 379
```

Db 320 DLGATTCDTLGMADVGTMCDFPRSCSVIEDDGLPSAFTTAHELGHVFNHMPHDNVKVCCE 379
QY 380 VFGKLRANHMSPPTLIQIDRANPWSACSAIIITDFLDGSHGDCLLDQPSKPISLPEDLP 439
Db 380 VFGKLRANHMSPPTLIQIDRANPWSACSAIIITDFLDGSHGDCLLDQPSKPISLPEDLP 439
QY 440 ASYTLSSQCELAFGVSGKPCPYMOCYCTKLWCTGKAKGQWVCOTRHPFADGTSCEGKLC 499
Db 440 ASYTLSSQCELAFGVSGKPCPYMOCYCTKLWCTGKAKGQWVCOTRHPFADGTSCEGKLC 499
QY 500 LKGACVERHNLNKHVRDGSWAKWDPYPCSRCTCGGQVQLARRQCTNPTPANGKYCEGVR 559
Db 500 LKGACVERHNLNKHVRDGSWAKWDPYPCSRCTCGGQVQLARRQCTNPTPANGKYCEGVR 559
QY 560 VKYRSCNLEPCPSSASGKSFREOQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLI 619
Db 560 VKYRSCNLEPCPSSASGKSFREOQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLI 619
QY 620 CRANGTGYYFVVLAPK-VVDGTLCSPDSTSVCGQKICAKGDCNGLGSKKRPDKGVCVCGD 678
Db 620 CRANGTGYYFVVLAPKVVVDGTLCSPDSTSVCGQKICAKGDCNGLGSKKRPDKGVCVCGD 679
QY 679 NKCKKVTGLFTKPMHGYNFVAIPAAGASSIDIRQRYKGLIGDDNYLALNKSQKYLIN 738
Db 680 NKCKKVTGLFTKPMHGYNFVAIPAAGASSIDIRQRYKGLIGDDNYLALNKSQKYLIN 739
QY 739 GHFVSAVERDLVKGSLRLRYSGTGFAVESLQASRPILFVEVLSVGMKTPPRVYSF 798
Db 740 GHFVSAVERDLVKGSLRLRYSGTGFAVESLQASRPILFVEVLSVGMKTPPRVYSF 799
QY 799 YLPKEPREDKSSHP 812
Db 800 YLPKEPREDKSSHP 813

RESULT 4

AAU72899

ID AAU72899 standard; Protein; 928 AA.

XX AAU72899;

AC AAU72899;

XX 26-FEB-2002 (first entry)

DT Human metalloprotease partial protein sequence #11.

DE Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;
KW vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;
KW hypertensive; neuroleptic; neuroprotective; anabolic;
KW anorectic; antiinflammatory; aspartyl protease; cysteine protease;
KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;
KW lung; prostrate; cervical; brain; ovarian; bladder; kidney; pain;
KW immune-related disease; cardiovascular disease; neuronal disease;
KW migraine; sexual dysfunction; mood disorder; attention disorder;
KW cognition disorder; hypertension; hypertention; psychotic disorder;
KW dyskinesia; metabolic disorder; inflammatory disorder.

XX Homo sapiens.

XX Wo200183782-A2.

XX 08-NOV-2001.

XX 04-MAY-2001; 2001WO-US14431.

XX 04-MAY-2000; 2000US-201879P.

XX (SUGEN-) SUGEN INC.

XX Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
PI Payne V;

XX WPI; 2002-041502/05.

XX N-PSDB; AA597182.

XX Novel protease polypeptide useful for screening for substances that may
PT be used to treat, e.g., cancers, immune-related diseases,
PT cardiovascular disease, migraine, pain, psychotic and inflammatory
PT disorders -
XX Claim 28; Figure 2G; 232pp; English.
PS The invention relates to an isolated, enriched, or purified protease
XX polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
CC screen for substances (S) that may modulate its activity. Administering
CC S (which modulates protease activity in vitro) may be used to treat a
CC disease or disorder selected from cancers (e.g., of tissues, of blood or
CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,
CC brain, ovarian, bladder or kidney), immune-related diseases and
CC disorders, cardiovascular disease, brain or neuronal-associated diseases
CC (e.g., central or peripheral nervous system diseases, migraine, pain,
CC sexual dysfunction, mood disorders, attention disorders, cognition
CC disorders, hypotension, hypertension, psychotic disorders, neurological
CC disorders and dyskinesias), metabolic disorders and inflammatory
CC disorders. (I) may also be useful as a diagnostic tool for a disease or
CC disorder such as those above. AAU72876-AAU72910 represent human
CC protease amino acid sequences of the invention.

XX Sequence 928 AA;

Query Match 66.8%; Score 635; DB 23; Length 928;

Best Local Similarity 99.9%; Pred No. 0;

Matches 755; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 20 EPEREVVPRLDPDINGRRYYWRGPDGSDGLIFQITAFQEDFYHLTDPDAFLAPAF 79

Db 45 EPEREVVPRLDPDINGRRYYWRGPDGSDGLIFQITAFQEDFYHLTDPDAFLAPAF 104

QY 80 STEHLGVPLQGLTGGSSDLRRCFYSGDYNAPDSPAAYSLGGLRGAFYGAEEVVISPL 139

Db 105 STEHLGVPLQGLTGGSSDLRRCFYSGDYNAPDSPAAYSLGGLRGAFYGAEEVVISPL 164

QY 140 PNASAPAAQRNSQGAHLQRRGVGGPGSDPTSRGVSAGWNPAILRALDPYKPRRAGFG 199

Db 165 PNASAPAAQRNSQGAHLQRRGVGGPGSDPTSRGVSAGWNPAILRALDPYKPRRAGFG 224

QY 200 ESRRRSRGRKRFYSIPRYVETLVVADBSMVKFGADLEHYLLTLATAARLVRHSIL 259

Db 225 ESRRRSRGRKRFYSIPRYVETLVVADBSMVKFGADLEHYLLTLATAARLVRHSIL 284

QY 260 NPINIVVVKVLLLRDRSDGPKVTGNAALTNRNFCWQKLNKVSQKHPYWDTAILFTRQ 319

Db 285 NPINIVVVKVLLLRDRSDGPKVTGNAALTNRNFCWQKLNKVSQKHPYWDTAILFTRQ 344

QY 320 DLGATTCDTLGMADVGTMCDFPRSCSVIEDDGLPSAFTTAHELGHVFNHMPHDNVKVCCE 379

Db 345 DLGATTCDTLGMADVGTMCDFPRSCSVIEDDGLPSAFTTAHELGHVFNHMPHDNVKVCCE 404

QY 380 VFGKLRANHMSPPTLIQIDRANPWSACSAIIITDFLDGSHGDCLLDQPSKPISLPEDLP 439

Db 405 VFGKLRANHMSPPTLIQIDRANPWSACSAIIITDFLDGSHGDCLLDQPSKPISLPEDLP 464

QY 440 ASYTLSSQCELAFGVSGKPCPYMOCYCTKLWCTGKAKGQWVCOTRHPFADGTSCEGKLC 499

Db 465 ASYTLSSQCELAFGVSGKPCPYMOCYCTKLWCTGKAKGQWVCOTRHPFADGTSCEGKLC 524

QY 500 LKGACVERHNLNKHVRDGSWAKWDPYPCSRCTCGGQVQLARRQCTNPTPANGKYCEGVR 559

Db 525 LKGACVERHNLNKHVRDGSWAKWDPYPCSRCTCGGQVQLARRQCTNPTPANGKYCEGVR 584

QY 560 VKYRSCNLEPCPSSASGKSFREOQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLI 619

Db 585 VKYRSCNLEPCPSSASGKSFREOQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLI 644

QY 620 CRANGTGYYFVVLAPK-VVDGTLCSPDSTSVCGQKICAKGDCNGLGSKKRPDKGVCVCGD 678

Db 645 CRANGTGYYFVVLAPKVVVDGTLCSPDSTSVCGQKICAKGDCNGLGSKKRPDKGVCVCGD 704


```
QY 679 NKCKKVTGLFTKPMHGYNFVAIPAGASSIDIRQYKGLIGDNDYILALKNKSGKYLLN 738
Db 705 NKCKKVTGLFTKPMHGYNFVAIPAGASSIDIRQYKGLIGDNDYILALKNKSGKYLLN 764
QY 739 GHFVSAVERDLVWKSLLRYSCTGTAVESLQASRP 774
Db 765 GHFVSAVERDLVWKSLLRYSCTGTAVESLQASRP 800

RESULT 5
AAE22542
ID AAE22542 standard; Protein; 367 AA.
XX
AC AAE22542;
XX
DT 26-JUL-2002 (first entry)
XX
DE Human protease #3.
XX
KW Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme.
XX
OS Homo sapiens.
XX
PN WO200226949-A2.
XX
PD 04-APR-2002.
XX
PF 27-SEP-2001; 2001WO-US30350.
XX
PR 29-SEP-2000; 2000US-236689P.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Friddle CJ, Hilbun E;
XX
DR WPI; 2002-372123/40.
XX
DR N-PSDB; AAD35570.
XX
PT Novel nucleic acid encoding a human protease, useful as a hybridization
PT probe for screening libraries and assessing gene expression patterns -
XX
PS Claim 7; Page 39-40; 41pp; English.
XX
CC The present sequence is novel human protein (NHP), human protease.
CC NHPs share structural similarity with animal proteases particularly
CC zinc metalloproteases. Sequences of the invention are useful in
CC therapeutic, diagnostic and pharmacogenomic applications. NHP
CC polynucleotides are used as hybridisation probes for screening
CC libraries and assessing gene expression patterns. They can also be
CC used for treating related biological disorders such as obesity, high
CC blood pressure, arthritis, connective tissue disorders and infertility.
CC They are also used in gene therapy.
XX
SQ Sequence 367 AA;
Query Match 36.3%; Score 345; DB 23; Length 367;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 EPEREVVPIRLDPDINGRRYWRGPDSDGGLIFQITAFQEDFYHLTPDAQFLAPAF 79
Db 20 EPEREVVPIRLDPDINGRRYWRGPDSDGGLIFQITAFQEDFYHLTPDAQFLAPAF 79
QY 80 STEHLGVPQLGTLGGSSDLRRCFYSGDVNAEPDSPAAYSLCGGLRGAFYGAEVVISPL 139
Db 80 STEHLGVPQLGTLGGSSDLRRCFYSGDVNAEPDSPAAYSLCGGLRGAFYGAEVVISPL 139
QY 140 PNASAPAAQNSOGAHLQRRGVPGSGDPTSRGCVASGNPAILLRALDPYKPRRAGFG 199
Db 140 PNASAPAAQNSOGAHLQRRGVPGSGDPTSRGCVASGNPAILLRALDPYKPRRAGFG 199
```

```
QY 200 ESRSRSSGRARFYSIPRYVETLVVADESMVKFGADLEHYLLTLAARLYRHPSIL 259
Db 200 ESRSRSSGRARFYSIPRYVETLVVADESMVKFGADLEHYLLTLAARLYRHPSIL 259
QY 260 NFINIVVVKVLLLRDRDGPVKVTGNAALTLRNFCAWOKKLNKVS DKHPEYWDTAILFTRQ 319
Db 260 NFINIVVVKVLLLRDRDGPVKVTGNAALTLRNFCAWOKKLNKVS DKHPEYWDTAILFTRQ 319
QY 320 DLGATTCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTAHEL 364
Db 320 DLGATTCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTAHEL 364

RESULT 6
AAE22540
ID AAE22540 standard; Protein; 321 AA.
XX
AC AAE22540;
XX
DT 26-JUL-2002 (first entry)
XX
DE Human protease #1.
XX
KW Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme.
XX
OS Homo sapiens.
XX
PN WO200226949-A2.
XX
PD 04-APR-2002.
XX
PF 27-SEP-2001; 2001WO-US30350.
XX
PR 29-SEP-2000; 2000US-236689P.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Friddle CJ, Hilbun E;
XX
DR WPI; 2002-372123/40.
XX
DR N-PSDB; AAD35568.
XX
PT Novel nucleic acid encoding a human protease, useful as a hybridization
PT probe for screening libraries and assessing gene expression patterns -
XX
PS Claim 5; Page 34-35; 41pp; English.
XX
CC The present sequence is novel human protein (NHP), human protease.
CC NHPs share structural similarity with animal proteases particularly
CC zinc metalloproteases. Sequences of the invention are useful in
CC therapeutic, diagnostic and pharmacogenomic applications. NHP
CC polynucleotides are used as hybridisation probes for screening
CC libraries and assessing gene expression patterns. They can also be
CC used for treating related biological disorders such as obesity, high
CC blood pressure, arthritis, connective tissue disorders and infertility.
CC They are also used in gene therapy.
XX
SQ Sequence 321 AA;
Query Match 31.6%; Score 300; DB 23; Length 321;
Best Local Similarity 100.0%; Pred. No. 1.6e-291;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 EPEREVVPIRLDPDINGRRYWRGPDSDGGLIFQITAFQEDFYHLTPDAQFLAPAF 79
Db 20 EPEREVVPIRLDPDINGRRYWRGPDSDGGLIFQITAFQEDFYHLTPDAQFLAPAF 79
QY 80 STEHLGVPQLGTLGGSSDLRRCFYSGDVNAEPDSPAAYSLCGGLRGAFYGAEVVISPL 139
Db 80 STEHLGVPQLGTLGGSSDLRRCFYSGDVNAEPDSPAAYSLCGGLRGAFYGAEVVISPL 139
```

CC	migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC	and autoimmune diseases. They can be used to treat patients afflicted
CC	with an invasive tumour, a brain tumour or brain injury.
XX	
SQ	Sequence 381 AA;
	Query Match 18.6%; Score 177; DB 21; Length 381;
	Best Local Similarity 100.0%; Pred. No. 3.1e-168;
	Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	255 HPSILNPINIVVVKLLLRDRDSGKVTGTGNAALTFLRNCAMQKKLNKVS DKHPEYWDTAI 314
Dd	16 HPSILNPINIVVVKLLLRDRDSGKVTGTGNAALTFLRNCAMQKKLNKVS DKHPEYWDTAI 75
QY	315 LFTFRDLGCATCTDITLGMADVGTCWCDPKRSCSVTEDDGLPSAFTTAAHELGHVFNNPHDNV 374
Dd	76 LFTFRDLGCATCTDITLGMADVGTCWCDPKRSCSVTEDDGLPSAFTTAAHELGHVFNNPHDNV 135
QY	375 KVCEEVFGKLRAHHMMSPTLIQIDRANPWSACSAAITDFLDSGHGDCILLDQPSKPI 431
Dd	136 KVCEEVFGKLRAHHMMSPTLIQIDRANPWSACSAAITDFLDSGHGDCILLDQPSKPI 192
	RESULT 8
AAB2157	
ID	AAB21257 standard; Protein; 505 AA.
XX	AAB21257;
DT	23-FEB-2001 (first entry)
DE	Rat metalloproteinase ADAMTS-5.
KW	Rat; ADAMTS-5; metalloproteinase; ADAM;
KW	a disintegrin and metalloproteinase domain; thrombospondin domain;
KW	vaccine; nontropic; neuroprotective; antiparkinsonian;
KW	cerebroprotective; cytostatic; antiarthritic; immunosuppressive;
KW	Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW	autoimmune disease; brain tumour; brain injury.
OS	Rattus norvegicus.
PX	WO200053774-A2.
PD	14-SEP-2000.
XX	
XX	08-MAR-2000; 2000WO-US06237.
XX	
PR	08-MAR-1999; 99US-0264585.
XX	(NEUR-) NEUROCRINE BIOSCIENCES INC.
PA	
PI	Keilner GS, Clark M, Maki RA;
XX	
DR	WFI; 2000-594326/56.
DR	N-PSDB; AAA95827.
XX	
PT	Polynucleotide encoding novel members of a disintegrin,
PT	metalloproteinase and thrombospondin domain protein family used to
PT	prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX	
PS	Claim 12; Fig 14; 129pp; English.
XX	
CC	The present sequence is rat metalloproteinase ADAMTS-5. The
CC	ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
CC	and Metalloproteinase Domain) family. Members of the ADAMTS family
CC	contain a thrombospondin domain in addition to the disintegrin and
CC	metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
CC	useful for the manufacture of medicaments for treating conditions
CC	associated with neuroinflammation and/or neurodegeneration, such as
CC	Alzheimer's disease, Parkinson's disease and stroke. They are also
CC	useful for treating conditions associated with cell proliferation, cell
CC	migration, inflammation and/or angiogenesis, such as cancer arthritis

CC and autoimmune diseases. They can be used to treat patients afflicted
CC with an invasive tumour, a brain tumour or brain injury.

XX Sequence 505 AA;

Query Match 16.2%; Score 154; DB 21; Length 505;

Best Local Similarity 100.0%; Pred. No. 4.3e-145;

Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 GPKVTGNAALTNRFCAROKKLNKVSDEKHPYWDTAILFTRODLGATTCOTLGMADVGT 337

Db 202 GPKVTGNAALTNRFCAROKKLNKVSDEKHPYWDTAILFTRODLGATTCOTLGMADVGT 261

QY 338 MCDPKRSCSVIEDGLPSAFTTAHELGHVFNPHDNKVKCEEVFGKLRANHMSPPTLIQI 397

Db 262 MCDPKRSCSVIEDGLPSAFTTAHELGHVFNPHDNKVKCEEVFGKLRANHMSPPTLIQI 321

QY 398 DRANPWSACSAIIITDPLDSGHGDCILLDQPSKPI 431

Db 322 DRANPWSACSAIIITDPLDSGHGDCILLDQPSKPI 355

RESULT 9

AAW78435

ID AAW78435 standard; Protein; 727 AA.

XX AC AAW78435;

XX 11-MAY-1999 (first entry)

XX Human ADAMTS-1 protein.

XX Metalloproteinase-disintegrin protein with a thrombospondin domain; ARDS;
KW ADAMTS-1; drug composition; foodstuff; leukocyte; thrombocyte; hepatitis;
KW blood; erythrocyte; inflammatory disease; rheumatoid arthritis; asthma;
KW nephritis; Crohn's disease; acute respiratory disease syndrome.

XX Homo sapiens.

XX WO9855643-A1.

XX 10-DEC-1998.

XX 03-JUN-1998; 98WO-JP02449.

XX 03-JUN-1997; 97JP-0160422.

XX (KURE) KUREHA CHEM IND CO LTD.

XX Hakezaki M, Hirose K, Inouchi E, Ishida Y, Ishioka K;

XX Kuno K, Matsushima K;

XX WPI: 1999-070277/06.

XX N-PSDB; AAX17990.

XX Human metalloproteinase-disintegrin protein with thrombospondin
PT domain - useful as leukocyte and thrombocyte decreasing and
PT erythrocyte increasing agent

XX Claim 1; Page 51-52; 82pp; Japanese.

XX This sequence represents a novel human metalloproteinase-disintegrin
CC protein with a thrombospondin domain (ADAMTS-1). The protein may be used
CC in drug compositions and foodstuffs, as an agent for decreasing the
CC leukocyte and thrombocyte blood count and increasing the erythrocyte
CC blood count, e.g. for treatment of inflammatory diseases such as
CC rheumatoid arthritis, hepatitis, nephritis, Crohn's disease, asthma
CC and ARDS.

XX Sequence 727 AA;

Query Match

Best Local Similarity 1.8%; Score 17; DB 20; Length 727;

Matches 100.0%; Pred. No. 1e-07;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372

Db 156 AFTTAHELGHVFNMPHD 172

RESULT 10

AAAB21265

ID AAB21265 standard; Protein; 896 AA.

XX AC AAB21265;

XX 23-FEB-2001 (first entry)

XX Mouse metalloproteinase ADAMTS-1.

XX Mouse; ADAMTS-1; metalloproteinase; ADAM;

XX a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; nontropic; neuroprotective; antiparkinsonian;
KW cerebroprotective; cytostatic; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.

XX Mus musculus.

XX WO200053774-A2.

XX 14-SEP-2000.

XX 08-MAR-2000; 2000WO-US06237.

XX 08-MAR-1999; 99US-0264585.

XX (NEUR-) NEUROCRINE BIOSCIENCES INC.

XX Kelner GS, Clark M, Maki RA;

XX WPI: 2000-594326/56.

XX Polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to
PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -

XX Disclosure; Fig 17; 129pp; English.

XX The present sequence is mouse metalloproteinase ADAMTS-1. The
CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
CC and Metalloproteinase Domain) family. Members of the ADAMTS family
CC contain a thrombospondin domain in addition to the disintegrin and
CC metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
CC useful for the manufacture of medicaments for treating conditions
CC associated with neuroinflammation and/or neurodegeneration, such as
CC Alzheimer's disease, Parkinson's disease and stroke. They are also
CC useful for treating conditions associated with cell proliferation, cell
CC migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC and autoimmune diseases. They can be used to treat patients afflicted
CC with an invasive tumour, a brain tumour or brain injury.

XX Sequence 896 AA;

Query Match 1.8%; Score 17; DB 21; Length 896;

Best Local Similarity 100.0%; Pred. No. 1.2e-07;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372

Db 380 AFTTAHELGHVFNMPHD 396

RESULT 11

AAAY49501

ID AAY49501 standard; Protein; 950 AA.

```

XX AC AAY49501;
XX DT 10-JAN-2000 (first entry)
XX DE Human METH1 protein.
XX KW Human: METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
XX KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
XX KW angiogenesis inhibitor; abnormal wound healing; inflammation;
XX KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
XX KW diabetic retinopathy; macula degeneration; haemangioma; detection;
XX KW arterial-venous malformation; immune deficiency.
XX OS Homo sapiens.
XX PN WO9937660-A1.
XX PD 29-JUL-1999.
XX PF 22-JAN-1999; 99WO-0501313.
XX PR 23-JAN-1998; 98US-0072298.
XX PR 28-AUG-1998; 98US-0098539.
XX PA (TRUE/) IRUELA-ARISPE L.
XX PA (HAST/) HASTINGS G A.
XX PA (RUBE/) RUBEN S M.
XX PI Irue-la-Arispe L, Hastings GA, Ruben SM;
XX DR WPI; 1999-590684/50.
XX DR N-PSDB; AA232000.
XX PS New Isolated metalloprotease thrombospondin polypeptides, useful for
XX PS treating hyperproliferative disorders, cancers or autoimmune disorders
XX PS Claim 10; Fig 1; 457pp; English.
XX CC AA232000 and AA232001 encode, and AAY49501 and AAY49502 represent, human
XX CC metalloprotease thrombospondin (METH) proteins METH1 and METH2
XX CC respectively. METH1 and METH2 have been found to be potent inhibitors of
XX CC angiogenesis both in vitro and in vivo. They can be used for treating
XX CC cancer and other disorders related to angiogenesis including abnormal
XX CC wound healing, inflammation, rheumatoid arthritis, psoriasis,
XX CC endometrial bleeding disorders, diabetic retinopathy, some forms of
XX CC macula degeneration, haemangiomas, and arterial-venous malformations.
XX CC They may be useful in treating deficiencies or disorders of the immune
XX CC system, by activating or inhibiting the proliferation, differentiation,
XX CC or mobilisation (chemotaxis) of immune cells. The etiology of these
XX CC immune deficiencies or disorders may be genetic, somatic, such as
XX CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
XX CC toxins), or infectious. They can also be used to treat inflammatory
XX CC conditions, both chronic and acute conditions. The products can also be
XX CC used for detection and diagnosis. AA232002 to AA232080, and AAY49503 to
XX CC AAY49511 represent sequences given in the exemplification of the present
XX CC invention.
XX SQ Sequence 950 AA;
Query Match 1.8%; Score 17; DB 20; Length 950;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 356 AFTTAHELGHVFNPHD 372
DB 379 AFTTAHELGHVFNPHD 395
|||||
RESULT 12
AAY53899
ID AAY53899 standard; Protein; 950 AA.

```

```

XX AC AAY53899;
XX DT 13-MAR-2000 (first entry)
XX DE Amino acid sequence of a murine ADAMTS-1 protein.
XX KW GON-1; metalloprotease; cell migration; modulator; ADAMTS-1;
XX KW metalloprotease domain; thrombospondin domain; abnormal cell migration;
XX KW organ shaping; sterility; cancer metastasis.
XX OS Mus sp.
XX PN WO9961656-A2.
XX PD 02-DEC-1999.
XX PF 28-MAY-1999; 99WO-US11918.
XX PR 29-MAY-1998; 98US-0087170.
XX PR 13-APR-1999; 99US-0129023.
XX PA (WISC ) WISCONSIN ALUMNI RES FOUND.
XX PI Kimble JE, Bieleloch RH;
XX DR WPI; 2000-072633/06.
XX PT Identifying modulators of proteins containing metalloprotease and
XX PT thrombospondin domains, potentially useful for controlling cell
XX PT migration and organ shaping -
XX PS Disclosure; Fig 1C; 60pp; English.
XX CC The present sequence represents a murine ADAMTS-1 protein. ADAMTS-1 is
XX CC a metalloprotease. The specification describes another related
XX CC metalloprotease, a C. elegans GON-1 protein, that lacks a transmembrane
XX CC domain and possesses a predicted metalloprotease domain between residues
XX CC 269-456. In C. elegans hermaphrodites, GON-1 is required for migration of
XX CC two distal tip cells to produce elongated tubes, whereas in males, GON-1
XX CC is required for migration of a single linker cell to produce a single
XX CC elongated tube. The protein is used in the method of the invention. The
XX CC specification describes a method for identifying a modulator of a
XX CC protein that contains a metalloprotease domain and a thrombospondin
XX CC domain. The method comprises treating a target organism, having a
XX CC developing gonadal cell that is responsive to the protein, with a test
XX CC compound, and determining any change in migration or shape of the cell
XX CC attributable to the test compound. The compounds identified are
XX CC potential therapeutic modulators of abnormal cell migration and organ
XX CC shaping, e.g. for rendering animals (specifically nematodes) sterile
XX CC and for inhibiting cancer metastases.
XX SQ Sequence 950 AA;
Query Match 1.8%; Score 17; DB 21; Length 950;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 356 AFTTAHELGHVFNPHD 372
DB 379 AFTTAHELGHVFNPHD 395
|||||
RESULT 13
AAB73549
ID AAB73549 standard; Protein; 950 AA.
XX AC AAB73549;
XX DT 07-AUG-2001 (first entry)
XX DE Human ADAM-type metalloprotease MDTs4, SEQ ID NO:4.

```

KW Human; MDTs4; ADAM-type metalloprotease; drug screening;
 KW A Disintegrin And Metalloprotease; cancer; arthritis.
 OS Homo sapiens.

XX JP2001017183-A.
 PN
 XX

PD 23-JAN-2001.

XX 09-JUL-1999; 99JP-0196584.

XX 09-JUL-1999; 99JP-0196584.

XX (YAMA) YAMANOUCHI PHARM CO LTD.

XX WPI; 2001-275950/29.

DR N-PSDB; AAH20224.

XX A new metal protease and its preparation for use as an anti-cancer and
 PT anti-arthritis therapeutic -

XX Claim 1; Page 12-14; 22pp; Japanese.

CC The invention relates to the novel human ADAM (A Disintegrin And
 CC Metalloprotease)-type metalloproteases MDTs4 (AAB73549) and MDTs5
 CC (AAB73550). The metalloproteases can be used for the treatment of
 CC cancers and arthritis. The invention also relates to the genes encoding
 CC MDTs4 and MDTs5, vectors and host cells containing the MDTs4 or MDTs5
 CC genes, the recombinant production of MDTs4 and MDTs5, and antibody
 CC specific for MDTs4 or MDTs5, and methods of screening for compounds
 CC which modulate the activity of MDTs4 and/or MDTs5. The present
 CC sequence represents human MDTs4.

XX Sequence 950 AA;

Query Match 1.8%; Score 17; DB 22; Length 950;

Best Local Similarity 100.0%; Pred. No. 1.3e-07;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 356 AFTTAHELGHVFNMPHD 372

Db 379 AFTTAHELGHVFNMPHD 395

RESULT 14

AAB50002

ID AAB50002 standard; Protein; 950 AA.

XX AAB50002;

DT 19-MAR-2001 (first entry)

DE Human METH1.

XX Human; METH1; metalloprotease; thrombospondin; angiogenesis inhibition;
 KW cancer therapy; benign tumour; ocular angiogenic disease;
 KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
 KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
 KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
 KW coronary collateral; cerebral collateral; arteriovenous malformation;
 KW ischaemic limb angiogenesis; Osler-Webber syndrome;
 KW plaque neovascularisation; telangiectasia; haemophilic joint;
 KW angiofibroma; fibromuscular dysplasia; wound granulation;
 KW Crohn's disease; atherosclerosis; birth control.

OS Homo sapiens.

XX WO200071577-A1.

XX 30-NOV-2000.

XX 25-MAY-2000; 2000WO-US14462.

PR 25-MAY-1999; 99US-0318208.
 PR 20-JUL-1999; 99US-0144882.
 PR 10-AUG-1999; 99US-0147823.
 PR 13-AUG-1999; 99US-0373658.
 PR 22-DEC-1999; 99US-0171503.
 PR 22-FEB-2000; 2000US-0183792.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA (TRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 PA (JONA/) JONAK Z L.
 PA (TRUL/) TRULLI S H.
 PA (FORN/) FORNWALD J A.
 PA (TERR/) TERRETT J A.

XX IrueLa-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
 PI Fornwald JA, Terrett JA;
 XX
 DR WPI; 2001-025136/03.
 DR N-PSDB; AAC90057.

XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
 PT inhibit angiogenesis in the treatment of disorders such as cancer,
 PT rheumatoid arthritis and psoriasis -
 XX Claim 15; Fig 1; 768pp; English.

XX The present sequence is human METH1 (ME for metalloprotease and TH for
 CC thrombospondin). METH1 can be used for inhibiting angiogenesis in an
 CC individual and for treating cancer, benign tumours, an ocular angiogenic
 CC disease, rheumatoid arthritis, psoriasis, delayed wound healing,
 CC endometriosis, vasculogenesis, granulations, hypertrophic scars, nonunion
 CC fractures, scleroderma, trachoma, vascular adhesions, myocardial
 CC angiogenesis, coronary collaterals, cerebral collaterals, arteriovenous
 CC malformations, ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or
 CC atherosclerosis. METH1 can also be used in birth control. METH1 can also
 CC be used in diagnostic methods for the prognosis of cancer.

XX Sequence 950 AA;

Query Match 1.8%; Score 17; DB 22; Length 950;

Best Local Similarity 100.0%; Pred. No. 1.3e-07;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 356 AFTTAHELGHVFNMPHD 372

Db 379 AFTTAHELGHVFNMPHD 395

RESULT 15

AAW80285

ID AAW80285 standard; Protein; 967 AA.

XX AAW80285;

XX 19-JAN-1999 (first entry)

XX Human integrin ligand polypeptide ITGL-TSP.

XX ITGL-TSP; integrin ligand; angiogenic disease; cancer; atherosclerosis;
 KW chronic inflammatory disorder; rheumatoid arthritis; tissue remodeling;
 KW macular degeneration; diabetic retinopathy; Alzheimer's disease; human;
 KW restenosis.

XX Homo sapiens.

XX EP874050-A2.

XX

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:20:53 ; Search time 24 Seconds
(without alignments)

1164.657 Million cell updates/sec

Title: US-10-009-332-1

Perfect score: 950

Sequence: 1 MLLIGILTLAGTAGGFE.....DQCNLHRKPOELDFCVLRPC 950

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2.6/ptodata/1/iaa/5A_COMB.pcp.*

2: /cgn2.6/ptodata/1/iaa/5B_COMB.pcp.*

3: /cgn2.6/ptodata/1/iaa/6A_COMB.pcp.*

4: /cgn2.6/ptodata/1/iaa/6B_COMB.pcp.*

5: /cgn2.6/ptodata/1/iaa/PCUUS_COMB.pcp.*

6: /cgn2.6/ptodata/1/iaa/backfiles.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	1.8	608	4	US-09-130-491-13
2	17	1.8	967	4	US-09-130-491-2
3	14	1.5	551	4	US-09-130-491-16
4	13	1.4	837	4	US-09-122-126B-2
5	13	1.4	874	4	US-09-369-364A-15
6	13	1.4	1882	4	US-09-369-364A-13
7	12	1.3	481	4	US-09-130-491-8
8	12	1.3	905	4	US-09-369-364A-9
9	11	1.2	930	4	US-09-369-364A-2
10	11	1.2	930	4	US-09-122-126B-15
11	10	1.1	518	4	US-09-369-364A-22
12	10	1.1	1205	4	US-09-491-522-11
13	10	1.1	1211	4	US-09-491-522-5
14	10	1.1	1224	4	US-09-930-872-4
15	9	0.9	859	4	US-09-369-364A-5
16	9	0.9	1081	4	US-09-369-364A-17
17	8	0.8	16	4	US-09-369-364A-23
18	8	0.8	17	4	US-09-122-126B-13
19	8	0.8	26	4	US-09-122-126B-4
20	8	0.8	109	4	US-09-199-637A-25
21	8	0.8	148	3	US-08-946-329A-60
22	8	0.8	525	4	US-09-369-364A-21
23	8	0.8	788	2	US-08-918-914-4
24	7	0.7	10	1	US-07-801-812A-6
25	7	0.7	10	1	US-08-487-568-6
26	7	0.7	12	5	PCT-US93-03748-11
27	7	0.7	13	1	US-07-801-812A-18

```

28 7 0.7 13 1 US-08-487-568-18 Sequence 18, Appl
29 7 0.7 13 4 US-08-871-561-29 Sequence 29, Appl
30 7 0.7 13 4 US-09-321-932B-29 Sequence 29, Appl
31 7 0.7 23 1 US-07-646-531D-7 Sequence 7, Appl
32 7 0.7 23 2 US-08-488-273-7 Sequence 7, Appl
33 7 0.7 23 4 US-09-197-770B-13 Sequence 13, Appl
34 7 0.7 23 6 5426100-7 Patent No. 5426100
35 7 0.7 28 2 US-08-449-933-9 Sequence 9, Appl
36 7 0.7 28 4 US-07-966-049A-9 Sequence 9, Appl
37 7 0.7 56 2 US-08-557-309B-29 Sequence 29, Appl
38 7 0.7 56 3 US-08-834-306-29 Sequence 29, Appl
39 7 0.7 56 4 US-08-993-674A-29 Sequence 29, Appl
40 7 0.7 56 4 US-09-256-976-29 Sequence 29, Appl
41 7 0.7 106 2 US-08-557-309B-44 Sequence 44, Appl
42 7 0.7 106 3 US-08-834-306-44 Sequence 44, Appl
43 7 0.7 106 4 US-08-993-674A-44 Sequence 44, Appl
44 7 0.7 106 4 US-09-256-976-44 Sequence 44, Appl
45 7 0.7 111 4 US-08-899-330-5 Sequence 5, Appl

```

ALIGNMENTS

```

RESULT 1
US-09-130-491-13
; Sequence 13, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-130-491-13

```

```

Query Match 1.8%; Score 17; DB 4; Length 608;
Best Local Similarity 100.0%; Pred. No. 9.4e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 356 AFTTAHELGHVFNMPHD 372
Db 91 AFTTAHELGHVFNMPHD 107

```

RESULT 2

```

US-09-130-491-2
; Sequence 2, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2

```

317 DTAI ETPODTC 323

; TITLE OF INVENTION: TANGO-71, TANGO 73, TANGO 74, TANGO 75, TANGO 76, AND TANGO 77


```

; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-130-491-8

```

```

Query Match 1.3%; Score 12; DB 4; Length 481;
Best Local Similarity 100.0%; Pred. No. 0.0047;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 603 WPKYSGVSPRD 614
    |||||
Db 202 WPKYSGVSPRD 213

```

```

RESULT 8
US-09-369-364A-9
; Sequence 9, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 905
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-8
US-09-369-364A-9

```

```

Query Match 1.3%; Score 12; DB 4; Length 905;
Best Local Similarity 100.0%; Pred. No. 0.0081;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 603 WPKYSGVSPRD 614
    |||||
Db 626 WPKYSGVSPRD 637

```

```

RESULT 9
US-09-369-364A-2
; Sequence 2, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 930
; TYPE: PRT
; ORGANISM: mus musculus ADAMTS-5
US-09-369-364A-2

```

```

Query Match 1.2%; Score 11; DB 4; Length 930;
Best Local Similarity 100.0%; Pred. No. 0.075;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 CDTLGMADVGT 337
    |||||
Db 376 CDTLGMADVGT 386

RESULT 10
US-09-122-126B-15
; Sequence 15, Application US/09122126B
; Patent No. 6451575
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: AGGREGAN DEGRADING METALLO PROTEASES
; FILE REFERENCE: DM6909
; CURRENT APPLICATION NUMBER: US/09/122,126B
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-122-126B-15

```

```

Query Match 1.2%; Score 11; DB 4; Length 930;
Best Local Similarity 100.0%; Pred. No. 0.075;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 327 CDTLGMADVGT 337
    |||||
Db 376 CDTLGMADVGT 386

```

```

RESULT 11
US-09-369-364A-22
; Sequence 22, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-5
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (99)
; OTHER INFORMATION: Xaa = M
US-09-369-364A-22

```

```

Query Match 1.1%; Score 10; DB 4; Length 518;
Best Local Similarity 100.0%; Pred. No. 0.41;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 670 DKCGVCGDN 679
    |||||
Db 307 DKCGVCGDN 316

```

```

RESULT 12
US-09-491-522-11
; Sequence 11, Application US/09491522
; Patent No. 6428998

```

```

; GENERAL INFORMATION:
; APPLICANT: Collige, Alain
; APPLICANT: Lapierre, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-491-522-11

Query Match 1.1%; Score 10; DB 4; Length 1205;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 DKCGVCGGDN 679
DB 704 DKCGVCGGDN 713
|||||

RESULT 13
US-09-491-522-5
; Sequence 5, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Collige, Alain
; APPLICANT: Lapierre, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows

```

```

; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1211 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-491-522-5

Query Match 1.1%; Score 10; DB 4; Length 1211;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 DKCGVCGGDN 679
DB 710 DKCGVCGGDN 719
|||||

RESULT 14
US-09-930-872-4
; Sequence 4, Application US/09930872
; Patent No. 6448388
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. 6448388 Human Proteases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0219-USA
; CURRENT APPLICATION NUMBER: US/09/930,872
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,852
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1224
; TYPE: PRT
; ORGANISM: homo sapiens
; US-09-930-872-4

Query Match 1.1%; Score 10; DB 4; Length 1224;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 527 PCSRTC GGCV 536
DB 597 PCSRTC GGCV 606
|||||

RESULT 15
US-09-369-364A-5
; Sequence 5, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00

```

```

; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 859
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-6
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (450)
; OTHER INFORMATION: Xaa = L
US-09-369-364A-5

```

```

Query Match      0.9%; Score 9; DB 4; Length 859;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTCGGV 536
    |||||
Db 522 CSRTCGGV 530

```

Search completed: April 29, 2003, 17:24:27
Job time : 26 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:25 ; Search time 35.9135 Seconds
(without alignments)
2738.218 Million cell updates/sec

Title: US-10-009-332-l_copy_213_950

Perfect score: 4043

Sequence: 1 FVSPRYVETLVVADESMWK.....DQCNLRKPOELDFCVLRPC 738

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_101002.*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4043	100.0	950	22 AAG62299	Human metalloprotease.
2	4043	100.0	950	23 AAE22541	Human protease #2.
3	3998	98.9	952	23 AAU74751	Human protease PRT
4	3743	5	92.6	23 AAU72899	Human metalloprotease
5	2260	55.9	727	20 AAU78435	Human ADAMTS-1 pro
6	2254	55.8	950	20 AAU49501	Human METH1 protei
7	2254	55.8	950	22 AAB73549	Human ADAM-type me
8	2254	55.8	950	22 AAB50002	Human METH1. Homo
9	2254	55.8	967	19 AAW80285	Human integrin lig
10	2254	55.8	967	20 AAY04142	Human Tango-71 pro

11	2254	55.8	968	22 AAB50011	Protein; SEQ ID 12
12	2241	55.4	967	20 AAW78189	Human secreted pro
13	2220	54.9	950	21 AAV53899	Amino acid sequenc
14	2036.5	51.9	896	21 AAB21365	Mouse metalloprote
15	1966	48.6	381	21 AAB21261	Human metalloprote
16	1920.5	47.5	505	21 AAB21257	Rat metalloprote
17	1911.5	47.3	905	22 AAB72284	Murine ADAMTS-8 am
18	1900	47.0	890	20 AAY49502	Human METH2 protei
19	1900	47.0	890	22 AAB50003	Human METH2. Homo
20	1891	46.8	889	22 AAB74946	Human ADAM type me
21	1888	46.7	680	21 AAB21251	Human metalloprote
22	1779	44.0	1629	23 AAG30703	Human aggrecanase
23	1779	44.0	1629	23 AAG14448	Human ADAMTS-SI pr
24	1779	44.0	1916	23 AAE19173	Human protease, PR
25	1779	44.0	1935	23 AAG72896	Human metalloprote
26	1774	43.9	1602	23 AAG30702	Human aggrecanase
27	1761	43.6	837	20 AAG75425	Human aggrecan deg
28	1761	43.6	837	22 AAG78228	Human metalloprote
29	1761	43.6	840	21 AAB21256	Human metalloprote
30	1760	43.5	837	21 AAY99429	Human PRO1563 (UNQ
31	1760	43.5	837	22 AAU29199	Human PRO polypept
32	1760	43.5	837	22 AAB66178	Protein of the inv
33	1757.5	43.5	1073	21 AAB21264	Human metalloprote
34	1757.5	43.5	1882	22 AAB72286	Human ADAMTS-9 am
35	1757.5	43.5	1934	22 AAB72301	Human ADAMTS-9 alt
36	1745	43.2	930	20 AAW75426	Human aggrecan deg
37	1740	43.0	929	21 AAB41226	Human aggrecan
38	1730.5	42.8	625	23 AAM48394	Rat aggrecanase.
39	1730.5	42.8	874	22 AAB72287	Murine ADAMTS-9 am
40	1726	42.7	930	22 AAB72280	Murine ADAMTS-5 am
41	1696	41.9	1907	23 AAU77133	Human metalloprote
42	1686	41.7	947	22 AAB86950	Human metalloprote
43	1685	41.7	958	21 AAB21255	Human metalloprote
44	1664	41.2	1505	23 AAB72897	Human metalloprote
45	1644.5	40.7	870	21 AAB21252	Rat metalloprote

ALIGNMENTS

RESULT 1
AAG62299
ID AAG62299 standard; protein; 950 AA.
XX AC
XX AAG62299;
DT 23-AUG-2001 (first entry)
XX
DE Human metalloprotease MDR56 protein.
XX
KW Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;
KW osteopathic; antiarthritic.
XX
OS Homo sapiens.
XX
XX WO200134785-A1.
XX
PD 17-MAY-2001.
XX
PF 10-NOV-2000; 2000WO-JP07917.
XX
PR 11-NOV-1999; 99JP-0321740.
PR 16-MAY-2000; 2000JP-0144020.
XX
XX (YAMA) YAMANOUCHI PHARM CO LTD.
XX (KAZU-) KAZUSA DNA RES INST.
PI Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
XX WPI; 2001-343602/36.
XX N-PSDB; AAH41003.
PT Metalloprotease with aggrecanase activity for treating joint diseases

PT especially osteoarthritis -
XX
PS Claim 1; Page 56-60; 85pp; Japanese.
XX
CC This invention relates to a metalloprotease with aggreganase activity.
CC The invention includes protein and DNA sequences of the metalloprotease,
CC vectors containing the DNA, host cells transformed by the vectors, and
CC antibodies directed against the metalloprotease. The antibodies, protein
CC and DNA sequences can be used in the treatment and prevention of joint
CC diseases, particularly osteoarthritis. The present invention may result in
CC osteopathic and antiarthritic activity. The present sequence represents
CC the metalloprotease of the invention termed MDT56.
XX
XX
SQ Sequence 950 AA;

Query Match 100.0%; Score 4043; DB 22; Length 950;
Best Local Similarity 100.0%; Pred. No. 4.2e-311;
Matches 738; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVSPRYVETLVVADESMVFKHGADLEHYLLTLTAARLYRHPSILNPINIVVVKVLL 60
Dy |||||
Dy 213 FVSPRYVETLVVADESMVFKHGADLEHYLLTLTAARLYRHPSILNPINIVVVKVLL 272
Qy 61 RDRSGPKVTGNAALTNRNFCAMQKLNKVS DKHPEYWDTAIFTRQDLCGATTCDTLGM 120
Dy |||||
Dy 273 RDRSGPKVTGNAALTNRNFCAMQKLNKVS DKHPEYWDTAIFTRQDLCGATTCDTLGM 332
Qy 121 ADVGTMCDPRKSCSVIEDDGLPSAFTTAHGLGHVFNPHDNVNVKCEVFGKLRANHMSP 180
Dy |||||
Dy 333 ADVGTMCDPRKSCSVIEDDGLPSAFTTAHGLGHVFNPHDNVNVKCEVFGKLRANHMSP 392
Qy 181 TLIQIDRANPWSACSAIITDLDGSHGDCLLDQSPKIPISLPEDLPASVYLSQOCELAF 240
Dy |||||
Dy 393 TLIQIDRANPWSACSAIITDLDGSHGDCLLDQSPKIPISLPEDLPASVYLSQOCELAF 452
Qy 241 GVSGKPCPYMOYCTKLWCTGKAKGMVYCOTRHPFADGTSCGEGKLCGLKACVERHNLNK 300
Dy |||||
Dy 453 GVSGKPCPYMOYCTKLWCTGKAKGMVYCOTRHPFADGTSCGEGKLCGLKACVERHNLNK 512
Qy 301 HRVDSGKAWKDPYGCSTCGGQVQLARRQCTNPTPANGKYCGVRYKYSCHNLEPCPS 360
Dy |||||
Dy 513 HRVDSGKAWKDPYGCSTCGGQVQLARRQCTNPTPANGKYCGVRYKYSCHNLEPCPS 572
Qy 361 SASGKSFREOCEAFNGYNHSTNRLTLAVAVVPKYSGVSPRDKKLCIRANGTGYFYVLA 420
Dy |||||
Dy 573 SASGKSFREOCEAFNGYNHSTNRLTLAVAVVPKYSGVSPRDKKLCIRANGTGYFYVLA 632
Qy 421 PKVVDGTLCSPDSTVCVQGCICAKGCDNGLSKRRDPKCGVCGDNKSKKVTGLFTKP 480
Dy |||||
Dy 633 PKVVDGTLCSPDSTVCVQGCICAKGCDNGLSKRRDPKCGVCGDNKSKKVTGLFTKP 692
Qy 481 MHGYNFVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQKYLNGHFVYSAVERDLV 540
Dy |||||
Dy 693 MHGYNFVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQKYLNGHFVYSAVERDLV 752
Qy 541 KGLLRYSGTGTAESLQASRPILFVLSVCKMTPPRVRYSYFLPKPREDKSSHP 600
Dy |||||
Dy 753 KGLLRYSGTGTAESLQASRPILFVLSVCKMTPPRVRYSYFLPKPREDKSSHP 812
Qy 601 KDPGRPSVLHNSVLSLSNQVEQDPRPARVAGSWGSPCSASCGLQKRAVDCRGSAQ 660
Dy |||||
Dy 813 KDPGRPSVLHNSVLSLSNQVEQDPRPARVAGSWGSPCSASCGLQKRAVDCRGSAQ 872
Qy 661 RTVPACDAAHRPVETQACEPCPTWELSAWSPCSKSGRGFORRLKCVGHGGRLLARDQ 720
Dy |||||
Dy 873 RTVPACDAAHRPVETQACEPCPTWELSAWSPCSKSGRGFORRLKCVGHGGRLLARDQ 932
Qy 721 CNLHRKPQELDFCVLRPC 738
Dy |||||
Dy 933 CNLHRKPQELDFCVLRPC 950

RESULT 2

AAE22541
ID AAE22541 standard; Protein; 950 AA.
XX
AC AAE22541;
XX
DT 26-JUL-2002 (first entry)
XX
DE Human protease #2.
XX
KW Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme.
XX
OS Homo sapiens.
XX
PN WO200226949-A2.
XX
PD 04-APR-2002.
XX
PF 27-SEP-2001; 2001WO-US30350.
XX
PR 29-SEP-2000; 2000US-236689P.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Friddle CJ, Hilbun E;
XX
PI WPI; 2002-372123/40.
XX
PI N-PSDB; AAD35569.
XX
PT Novel nucleic acid encoding a human protease, useful as a hybridization
PT probe for screening libraries and assessing gene expression patterns -
XX
PS Claim 6; Page 36-38; 4lpp; English.
XX
CC The present sequence is novel human protein (NHP), human protease.
CC NHPs share structural similarity with animal proteases particularly
CC zinc metalloproteases. Sequences of the invention are useful in
CC therapeutic, diagnostic and pharmacogenomic applications. NHP
CC polynucleotides are used as hybridisation probes for screening
CC libraries and assessing gene expression patterns. They can also be
CC used for treating related biological disorders such as obesity, high
CC blood pressure, arthritis, connective tissue disorders and infertility.
CC They are also used in gene therapy.
XX
SQ Sequence 950 AA;

Query Match 100.0%; Score 4043; DB 23; Length 950;
Best Local Similarity 100.0%; Pred. No. 4.2e-311;
Matches 738; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FVSPRYVETLVVADESMVFKHGADLEHYLLTLTAARLYRHPSILNPINIVVVKVLL 60
Dy |||||
Dy 213 FVSPRYVETLVVADESMVFKHGADLEHYLLTLTAARLYRHPSILNPINIVVVKVLL 272
Qy 61 RDRSGPKVTGNAALTNRNFCAMQKLNKVS DKHPEYWDTAIFTRQDLCGATTCDTLGM 120
Dy |||||
Dy 273 RDRSGPKVTGNAALTNRNFCAMQKLNKVS DKHPEYWDTAIFTRQDLCGATTCDTLGM 332
Qy 121 ADVGTMCDPRKSCSVIEDDGLPSAFTTAHGLGHVFNPHDNVNVKCEVFGKLRANHMSP 180
Dy |||||
Dy 333 ADVGTMCDPRKSCSVIEDDGLPSAFTTAHGLGHVFNPHDNVNVKCEVFGKLRANHMSP 392
Qy 181 TLIQIDRANPWSACSAIITDLDGSHGDCLLDQSPKIPISLPEDLPASVYLSQOCELAF 240
Dy |||||
Dy 393 TLIQIDRANPWSACSAIITDLDGSHGDCLLDQSPKIPISLPEDLPASVYLSQOCELAF 452
Qy 241 GVSGKPCPYMOYCTKLWCTGKAKGMVYCOTRHPFADGTSCGEGKLCGLKACVERHNLNK 300
Dy |||||
Dy 453 GVSGKPCPYMOYCTKLWCTGKAKGMVYCOTRHPFADGTSCGEGKLCGLKACVERHNLNK 512
Qy 301 HRVDSGKAWKDPYGCSTCGGQVQLARRQCTNPTPANGKYCGVRYKYSCHNLEPCPS 360
Dy |||||

Db 513 HRVDSWAKWDPYGPCSRTCCGGVQLARRQCTNPTPANGKYCEGVYKVRKSCNLEPCPS 572
QY 361 SASGKSFREEOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLCIRANGTGYFYVLA 420
Db 573 SASGKSFREEOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLCIRANGTGYFYVLA 632
QY 421 PKVVDGTLCSPDSTSVQVQKCIKAGCDGNLGSKKRFDKCGVCGDNKSKKVTGLFTKP 480
Db 633 PKVVDGTLCSPDSTSVQVQKCIKAGCDGNLGSKKRFDKCGVCGDNKSKKVTGLFTKP 692
QY 481 MHGYNFVVAIPAGASSIDIRQGYKGLIGDDNYLALNSQKYLNGHFVYSAVERDLV 540
Db 693 MHGYNFVVAIPAGASSIDIRQGYKGLIGDDNYLALNSQKYLNGHFVYSAVERDLV 752
QY 541 KGSLLRYSGTGAVESLQASRPILPTVEVLVSVKMTTPRVRYSFVLPKREPREDKSSHP 600
Db 753 KGSLLRYSGTGAVESLQASRPILPTVEVLVSVKMTTPRVRYSFVLPKREPREDKSSHP 812
QY 601 KDPGRPSVLHNSVLSLNQVEQDDRRPARWAGSWGPCSSASCGSLQKRAVDCRGSAGQ 660
Db 813 KDPGRPSVLHNSVLSLNQVEQDDRRPARWAGSWGPCSSASCGSLQKRAVDCRGSAGQ 872
QY 661 RTVPACDAHRPVTQACGEPCTWELSAWSPCKSCGGRGFORSLKCVCHGGRLLARDQ 720
Db 873 RTVPACDAHRPVTQACGEPCTWELSAWSPCKSCGGRGFORSLKCVCHGGRLLARDQ 932
QY 721 CNLHRKPOELDFCVLRPC 738
Db 933 CNLHRKPOELDFCVLRPC 950

RESULT 3
AAU74751
ID AAU74751 standard; Protein: 952 AA.
AC AAU74751;
XX
XX
XX
XX 09-APR-2002 (first entry)
XX
DE Human protease PRts-11 protein sequence.
XX Human; protease; PRts; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis.
XX
OS Homo sapiens.
XX
XX W0200198468-A2.
XX
PD 27-DEC-2001.
XX
XX 13-JUN-2001; 2001WO-US19178.
XX
XX 16-JUN-2000; 2000US-212336P.
XX 22-JUN-2000; 2000US-213955P.
XX 29-JUN-2000; 2000US-215396P.
XX 07-JUL-2000; 2000US-216821P.
XX 14-JUL-2000; 2000US-218946P.
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Walla NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;
PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Kallick DA;
XX
XX WPI: 2002-090437/12.
XX N-PSDB; ABK12894.

PT Twenty one human proteases (referred to as PRts-1 to PRts-21), useful
PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
PT proliferative (e.g. cancer) disorders -
XX
PS Claim 1; Page 144-146; 177pp; English.
XX
XX The present invention relates to twenty one new human proteases,
CC referred to as PRts-1 to PRts-21. The PRts polynucleotides and
CC polypeptides of the invention are useful in the diagnosis, treatment and
CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
CC myocardial infarction, autoimmune/inflammatory e.g. acquired
CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
CC endometriosis disorders. Numerous other examples of each disorder are
CC given in the specification. The present protein sequence represents
CC the human protease PRts-11 protein of the invention.

Sequence 952 AA;

Query Match 98.9%; Score 3998; DB 23; Length 952;
Best Local Similarity 99.3%; Pred No. 1.5e-307;
Matches 735; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLLAARLYRHPHSILNPINIVVVKVLL 60
Db 213 FVSIPRYVETLVVADESVMKFGADLEHYLLTLLAARLYRHPHSILNPINIVVVKVLL 272
QY 61 RDRDSGPKVTGNAALTNRNFCWOKKLNKVSXKHPHYWDTAILFTRODLGATTCDDTLGM 120
Db 273 RDRDSGPKVTGNAALTNRNFCWOKKLNKVSXKHPHYWDTAILFTRODLGATTCDDTLGM 332
QY 121 ADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVNMFDNPKVCEVEFGKLRANHMSP 180
Db 333 ADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVNMFDNPKVCEVEFGKLRANHMSP 392
QY 181 TLIIQIDRANPWSACSAIIITDPLDSHGDCLLDQSPKPSILPEDLPASAYTLSQOCELA 240
Db 393 TLIIQIDRANPWSACSAIIITDPLDSHGDCLLDQSPKPSILPEDLPASAYTLSQOCELA 452
QY 241 GVGSKPCPYMOYCTKLWCTGKAKGMVQTRHFPWADGTSCGEGKLCCLKGACVERHNLK 300
Db 453 GVGSKPCPYMOYCTKLWCTGKAKGMVQTRHFPWADGTSCGEGKLCCLKGACVERHNLK 512
QY 301 HRVDGSWAKWDPYGPCSRTCCGGVQLARRQCTNPTPANGKYCEGVYKVRKSCNLEPCPS 360
Db 513 HRVDGSWAKWDPYGPCSRTCCGGVQLARRQCTNPTPANGKYCEGVYKVRKSCNLEPCPS 572
QY 361 SASGKSFREEOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLCIRANGTGYFYVLA 420
Db 573 SASGKSFREEOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLCIRANGTGYFYVLA 632
QY 421 PK-VVDGTLCSPDSTSVQVQKCIKAGCDGNLGSKKRFDKCGVCGDNKSKKVTGLFTK 479
Db 633 PKVVDGTLCSPDSTSVQVQKCIKAGCDGNLGSKKRFDKCGVCGDNKSKKVTGLFTK 692
QY 480 MHGYNFVVAIPAGASSIDIRQGYKGLIGDDNYLALNSQKYLNGHFVYSAVERDLV 539
Db 693 MHGYNFVVAIPAGASSIDIRQGYKGLIGDDNYLALNSQKYLNGHFVYSAVERDLV 752
QY 540 VKGSLLRYSGTGAVESLQASRPILPTVEVLVSVKMTTPRVRYSFVLPKREPREDKSSH 599
Db 753 VKGSLLRYSGTGAVESLQASRPILPTVEVLVSVKMTTPRVRYSFVLPKREPREDKSSH 812
QY 600 PKDPR-GPSVLHNSVLSLNQVEQDDRRPARWAGSWGPCSSASCGSLQKRAVDCRGS 658
Db 813 PPHPRGPSVLHNSVLSLNQVEQDDRRPARWAGSWGPCSSASCGSLQKRAVDCRGS 872
QY 659 GQRTVPACDAHRPVTQACGEPCTWELSAWSPCKSCGGRGFORSLKCVCHGGRLLAR 718

Db	873	Q0TVFACDAHHPVETQAGPECP2WELSAWSPCKSCGRGFORSLKCVCHGRLLAR	932
QY	719	DQCNLHRKPQELDFCVLRPC	738
Db	933	DQCNLHRKPQELDFCVLRPC	952
RESULT 4			
AAU72899			
ID	AAU72899	standard; Protein; 928 AA.	
XX	AAU72899;		
XX	26-FEB-2002	(first entry)	
XX	Human metalloprotease partial protein sequence #11.		
XX	Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;		
KW	vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser		
KW	hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;		
KW	anorectic; antiinflammatory; aspartyl protease; cysteine protease;		
KW	metalloprotease; serine protease; cancer; haematopoietic; breast; colon;		
KW	lung; prostrate; cervical; brain; ovarian; bladder; kidney; pain;		
KW	immune-related disease; cardiovascular disease; neuronal disease;		
KW	migraine; sexual dysfunction; mood disorder; attention disorder;		
KW	cognition disorder; hypotension; hypertension; psychotic disorder;		
KW	dykinesia; metabolic disorder; inflammatory disorder.		
OS	Homo sapiens.		
XX	W0200183782-A2.		
XX	08-NOV-2001.		
XX	04-MAY-2001; 2001WO-US14431.		
XX	04-MAY-2000; 2000US-201879P.		
XX	(SUGE-) SUGEN INC.		
XX	Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;		
PI	Payne V;		
XX	WPI; 2002-041502/05.		
DR	N-PSDB; RAS97182.		
XX	Novel protease polypeptide useful for screening for substances that may		
PT	be used to treat, e.g., cancers, immune-related diseases,		
PT	cardiovascular disease, migraine, pain, psychotic and inflammatory		
PT	disorders -		
XX	Claim 28; Figure 2G; 232pp; English.		
XX	The invention relates to an isolated, enriched, or purified protease		
XX	polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to		
CC	screen for substances (S) that may modulate its activity. Administering		
CC	S (which modulates protease activity in vitro) may be used to treat a		
CC	disease or disorder selected from cancers (e.g., of tissues, of blood or		
CC	haematopoietic origin, of the breast, colon, lung, prostate, cervical,		
CC	brain, ovarian, bladder or kidney), immune-related diseases and		
CC	disorders, cardiovascular disease, brain or neuronal-associated diseases		
CC	(e.g., central or peripheral nervous system diseases, migraine, pain,		
CC	sexual dysfunction, mood disorders, attention disorders, cognition		
CC	disorders, hypotension, hypertension, psychotic disorders, neurological		
CC	disorders and dyskinesias), metabolic disorders and inflammatory		
CC	disorders. (I) may also be useful as a diagnostic tool for a disease or		
CC	disorder such as those above. AAU72876-AAU72910 represent human		
CC	protease amino acid sequences of the invention.		
XX	Sequence 928 AA;		
SQ	Query Match 92.6%; Score 3743.5; DB 23; Length 928;		
	Best Local Similarity 93.4%; Pred. No. 2.1e-287;		

Db 714 VTSAPGCHDIIITPTGATNIEVKQNRGRNNGSFLAIAKADGTYLLNGDYTLSTLEQ 773
 QY 537 DLVVKGSLLRYSGTGTAVERSLQASRPILPTVEVLVSGKMTPPRVYSFYLPKEPREDK 596
 Db 774 DIMKGVVLRYSGSSAALERIRSFSLKPLETIQVLTGVNALRPKIKYTFVKKK---K 829
 QY 597 SSHPKDPRGSPVLSLNSVLSNOVEQDDPPARWAGSWGPGCSASGSGLOKRAVDORG 656
 Db 830 ES-----FNAIPTFS-----AWIEEWGECSKCELGWQRRLECDR 866
 QY 657 SAGQRTVPACDAH--RPVETOACGE-PCPTWELSAWSPCSKSGRGFORSLKCVGHGG 713
 Db 867 INGO---PASECAKEVKPASTRPCADHPCPQWLGWSSCSKTGGRYKRSILKCLSHDG 923
 QY 714 RLRLAQCNLHRKPQE-LDFCVLRPC 738
 Db 924 GVLSHESCDPLKKPKHFIDECTWAC 949

RESULT 8
 AAB50002
 ID AAB50002 standard; Protein; 950 AA.
 AC AAB50002;
 XX
 DT 19-MAR-2001 (first entry)
 XX
 DE Human METH1.
 XX
 KW Human; METH1; metalloprotease; thrombospondin; angiogenesis inhibition;
 KW cancer therapy; benign tumour; ocular angiogenic disease;
 KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
 KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
 KW scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;
 KW coronary collateral; cerebral collateral; arteriovenous malformation;
 KW ischaemic limb angiogenesis; Osler-Webber syndrome;
 KW plaque neovascularisation; telangiectasia; haemophilic joint;
 KW angiofibroma; fibromuscular dysplasia; wound granulation;
 KW Crohn's disease; atherosclerosis; birth control.
 XX
 OS Homo sapiens.
 XX
 PN WO200071577-A1.
 PD 30-NOV-2000.
 XX
 PF 25-MAY-2000; 2000WO-US14462.
 XX
 PR 25-MAY-1999; 99US-0318208.
 PR 20-JUL-1999; 99US-0144882.
 PR 10-AUG-1999; 99US-0147823.
 PR 13-AUG-1999; 99US-0373658.
 PR 22-DEC-1999; 99US-0171503.
 PR 22-FEB-2000; 2000US-0183792.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA (IRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 PA (JONA/) JONAK Z L.
 PA (TRUL/) TRULLI S H.
 PA (FORN/) FORNWALD J A.
 PA (TERR/) TERRETT J A.
 XX
 PA IrueLA-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
 PI Fornwald JA, Terrett JA;
 XX
 DR WPI: 2001-025136/03.
 DR N-FSDB: AAC90057.
 XX
 PT METH1 and METH2 polynucleotides and encoded polypeptides, used to

PT inhibit angiogenesis in the treatment of disorders such as cancer,
 PT rheumatoid arthritis and psoriasis -
 PS Claim 15; Fig 1; 768pp; English.
 CC The present sequence is human METH1 (ME for metalloprotease and TH for
 CC thrombospondin). METH1 can be used for inhibiting angiogenesis in an
 CC individual, and for treating cancer, benign tumours, an ocular angiogenic
 CC disease, rheumatoid arthritis, psoriasis, delayed wound healing, nonunion
 CC endometriosis, vasculogenesis, granulations, hypertrophic scars, nonunion
 CC fractures, scleroderma, trachoma, vascular adhesions, myocardial
 CC angiogenesis, coronary collaterals, cerebral collaterals, arteriovenous
 CC malformations, ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma, plaque
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or
 CC atherosclerosis. METH1 can also be used in birth control. METH1 can also
 CC be used in diagnostic methods for the prognosis of cancer.
 XX

SQ Sequence 950 AA;

Query Match 55.8%; Score 2254; DB 22; Length 950;
 Best Local Similarity 54.0%; Pred. No. 1.8e-169;
 Matches 403; Conservative 131; Mismatches 172; Indels 40; Gaps 11;

QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATARLYRHPHSILNPINIVVVKVLL 60
 Db 236 FVSSHRVYVETMLVADQSMAEFHGSLKHYLLTLFSAARLYKHPSIRNSVLSVVKILVI 295
 QY 61 RDRSGPKVTGNAALTLRNFCAWOKLNKVDKHPYWDYTAILFTRODLGATTCDTLGM 120
 Db 296 HDEQKGPEVTSNAALTLRNFCNMQKQNPSPDRDAEHYDTALTFTRODLGSGTCDTLGM 355
 QY 121 ADVGTMCDPKRSCVIEDGLPSAFTTAHELGHVFNPHDNKVKCEVEFKLRANHMSP 180
 Db 356 ADVGTVCDPSSCSVIEDGLQAAFTTAHELGHVFNPHDDAKCASLNGVNDSHMMAS 415
 QY 181 TLIQIDRANPWSACSAIITDLDGSHGDCILDDQSPISLPEDLPASVTLSSOCELAF 240
 Db 416 MSLNLDHSQWSPCSAYMTISFLDNGHGECLMDKPNPQIPLGDLPTSDANHCQCFTF 475
 QY 241 GVGSKPCP-TMQYCTKLWCTGKAKQMVCTRHFPADGTSCGEGKLCGLKACVVERHNLN 299
 Db 476 GEDSKHCPDAASTCTLWCTGSGVLVCQTKHPPADGTSCGEGKWCINGKCVNKTID-R 534
 QY 300 KH---RYDGSNAKWDPYPCSRCTCGGVQVLARQCTNPTPANGKCYCEGVRYRSCNLE 356
 Db 535 KHEDTPFHGSGWGMGMDCSRCTCGGVQVTMRCDNPVKNKGKCYCEGKRVYRSCNLE 594
 QY 357 PCSSASGKSFREBQCEAFNGYNHSTNRLTLAVAWVPKYSVSGVSPDRCKLICRANGTYF 416
 Db 595 DCPDN-NGKTFREBQCEAHNEFSKASFGSGPAVEWIPKYAGVSPKORCKLICQAKGIGYF 653
 QY 417 YVLAPKVVDTGLCSPDSTSVCGVKCIKACDGNLGSKKRFDKCGVCGGDNKSKCKVTGL 476
 Db 654 FVLQPKVVDGTPCSPDSTSVCGVQCVKAGCDRIIDSKKFKDCGCGGNGSTCKTISGS 713
 QY 477 FTKPMHGYFNFWALPAGASSIDIRQRYGKGLIGDNNYLAALKNSOGKYLNGHGVVSAVER 536
 Db 714 VTSAPGCHDIIITPTGATNIEVKQNRGRNNGSFLAIAKADGTYLLNGDYTLSTLEQ 773
 QY 537 DLVVKGSLLRYSGTGTAVERSLQASRPILPTVEVLVSGKMTPPRVYSFYLPKEPREDK 596
 Db 774 DIMKGVVLRYSGSSAALERIRSFSLKPLETIQVLTGVNALRPKIKYTFVKKK---K 829
 QY 597 SSHPKDPRGSPVLSLNSVLSNOVEQDDPPARWAGSWGPGCSASGSGLOKRAVDORG 656
 Db 830 ES-----FNAIPTFS-----AWIEEWGECSKCELGWQRRLECDR 866
 QY 657 SAGQRTVPACDAH--RPVETOACGE-PCPTWELSAWSPCSKSGRGFORSLKCVGHGG 713
 Db 867 INGO---PASECAKEVKPASTRPCADHPCPQWLGWSSCSKTGGRYKRSILKCLSHDG 923
 QY 714 RLRLAQCNLHRKPQE-LDFCVLRPC 738

XX PS Claim 8: Fig 1; 84pp; English.

XX CC The present sequence represents human Tango-71. Tango polypeptides are

CC useful for identifying compounds which bind the polypeptide via direct

CC binding, competition binding assays or Tango-71, -73, -74, 76 or -83

CC mediated signal transduction. Tango polypeptides are also useful for

CC identifying modulating compounds by determining effect on Tango activity.

CC Tango polypeptides and nucleic acids are useful for diagnosing diseases

CC related to aberrant expression of Tango, and Tango polypeptides are

CC useful for raising antibodies which can be used in diagnostic assays for

CC detection of Tango, and also for generating anti-idiotype antibodies for

CC prevention and protection.

XX SQ Sequence 967 AA;

Query Match 55.8%; Score 2254; DB 20; Length 967;

Best Local Similarity 54.0%; Pred. No. 1.8e-169;

Matches 403; Conservative 131; Mismatches 172; Indels 40; Gaps 11;

QY 1 FVSIPRYVETLVVADESMVKFHGADLEHYLLTLLATAARLYRHPSILNPINIVVKKVLL 60

DB 253 FVSSHRYVETMLVADQSAEFGHSGKHYLLTFLSVAARLYKHPISNVSILVVKILVI 312

QY 61 RDRDSGPKVTGNAALTILNFCQAWKKLNKYSKHPYWDYTAILFTRODLGATCDTIGM 120

DB 313 HDEQKGEVTSNAALTILNFCQAWKKLNKYSKHPYWDYTAILFTRODLGATCDTIGM 372

QY 121 ADVGTMCDPKRCSCVIEDDGLPSAFTTAHELGHVFNMPDNVKEEVEFGKLRANHMSP 180

DB 373 ADVGTVCDFPSCSVIEDDGLQAETTAHELGHVFNMPDHAQKASLNGVQNDSHMMAS 432

QY 181 TLIIQIDRANPWSACSAIITFDLSGHGDCLLDQPSKIPSLPEDLPGASYTLSQOCELAF 240

DB 433 MSLNLDHQPSPSCSAYMITSFELNGHGECLMDKPNQIPLPGDLPGTSDANRQOQFTF 492

QY 241 GVSGKPCP-QYMCYCKLACTKAGQYVQTRHFPWADGTSCGSEKCLGKACVERNNLN 299

DB 493 GEDSKHCPDAASTCLTSLTSGVLVCQTKHFPWADGTSCGSEKCLGKACVERNNLN 551

QY 300 KH--RVDGSAKWPDPYPCSTCGGCVOLARROCTNTPANGKYGCVGVKYSRNL 356

DB 552 KHFDTPFHGSGMGWPGWDCSTCGGCVQYTWRECDNVPKNGKYGCVGVKYSRNL 611

QY 357 PCPSASGSKFPEOCEAFNGYHNSNRLTLAVAWPKYSGVSPDRCKLICRANGTYG 416

DB 612 DCPDN-NGKTFREOCEAHNEFSKASFGSGPAVEWPKYAGVSPKDRCKLICQAKGIGYF 670

QY 417 VYLAQKVVDTLCPDSTSVQVQKCIKAGCDNLGSKKRFDCGVCVGGDNKSKKVTGL 476

DB 671 FVLQPKVVDGTFPCPDSTSVQVQKCIKAGCDNLGSKKRFDCGVCVGGDNKSKKVTGL 730

QY 477 FTKPMHGYNVFAIPAGASSIDIRQYKYLIGDDNLALKNSSOGKYLNGHFVVSVAVER 536

DB 731 VTSAPKGYHDIITPTGATNIEVKORNGSRNNGSFLAIAADGTILNGDYTLSTLEQ 790

QY 537 DLVVKGSLRLSGTGCTAVESQARPLEPLTVLSVGMKTPPRVRYSPFLKPEPREDK 596

DB 791 DIMYKGVLYRYSGSAALERTSRFSLKEPLTIQVLTWGNALRPKIKYTFVKKK---K 846

QY 597 SSHPKDPGRPSVLNLSVLSNOVEQDDRPAPRWAGSWGPCASGSLGKRAVDCRG 656

DB 847 ES-----FNAIPTFS-----AMVIEEWGECSSKSELGQWRRLVECRD 883

QY 657 SAGQRTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCSKSGRGFORSLKVCVGHG 713

DB 884 INGO---PASECAKEVPASTRPCADHPCPOWQJLGEWSSCKTKCGKIKRSLKCLSHDG 940

QY 714 RLLAQOQNLHRKPOE-LDFCVLRPC 738

DB 941 GVLSHESCDPLKPKHFDICTMAEC 966

RESULT 11

AAB50011

ID AAB50011 standard; Protein; 968 AA.

XX AC AAB50011;

XX DT 19-MAR-2001 (first entry)

XX DE Protein; SEQ ID 125.

XX KW Human; METH1; metalloprotease; thrombospondin; angiogenesis inhibition; cancer therapy; benign tumour; ocular angiogenic disease;

XX KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;

XX KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;

XX KW scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;

XX KW coronary collateral; cerebral collateral; arteriovenous malformation;

XX KW ischaemic limb angiogenesis; Osler-Webber syndrome;

XX KW plaque neovascularisation; telangiectasia; haemophilic joint;

XX KW angiofibroma; fibromuscular dysplasia; wound granulation;

XX KW Crohn's disease; atherosclerosis; birth control.

OS Homo sapiens.

XX WO200071577-A1.

XX PD 30-NOV-2000.

XX PF 25-MAY-2000; 2000WO-US14462.

XX PR 25-MAY-1999; 99US-0318208.

XX PR 20-JUL-1999; 99US-0144882.

XX PR 10-AUG-1999; 99US-0147823.

XX PR 13-AUG-1999; 99US-0373658.

XX PR 22-DEC-1999; 99US-0171503.

XX PR 22-FEB-2000; 2000US-0183792.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX (IRUE/) IRUELA-ARISPE L.

XX (HAST/) HASTINGS G A.

XX (RUBE/) RUBEN S M.

XX (JONA/) JONAK Z L.

XX (TRUL/) TRULLI S H.

XX (FORN/) FORNWALD J A.

XX (TERE/) TERRETT J A.

XX IrueLa-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH; Fornwald JA, Terrett JA;

XX WPI; 2001-025136/03.

XX METH1 and METH2 polynucleotides and encoded polypeptides, used to inhibit angiogenesis in the treatment of disorders such as cancer, rheumatoid arthritis and psoriasis -

XX Claim 15; Pages 759-763; 768pp; English.

XX The present invention relates to human METH1 and METH2 (ME for metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).

XX METH can be used for inhibiting angiogenesis in an individual, and for treating cancer, benign tumours, an ocular angiogenic disease, rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis, vasculogenesis, granulations, hypertrophic scars, nonunion fractures, scleroderma, trachoma, vascular adhesions, myocardial angiogenesis, coronary collaterals, cerebral collaterals, arteriovenous malformations, ischaemic limb angiogenesis, Osler-Webber syndrome, plaque neovascularisation, telangiectasia, haemophilic joints, angiofibroma, fibromuscular dysplasia, wound granulation, Crohn's disease or atherosclerosis. METH can also be used in birth control. METH can also be used in diagnostic methods for the prognosis of cancer. The present sequence is a protein isolated in the present invention.

XX

Query Match	55.8%;	Score	2254;	DB	22;	Length	968;
Best Local Similarity	54.0%;	Pred.	No. 1.8e-169;				
Matches	403;	Conservative	131;	Mismatches	172;	Indels	40;
Gaps	11;						

QY	1	FVSIPIRYVETLVVADESVMVKHGADLEHYLTLLTATAAARLYRHPISILNPINIVVVKVILL	60
DB	254	FVSSHRYVETLMVADQSMAEFHGSLKHYLETLFESVAARLYKHPSIRNSVSLVVVKILVI	313
QY	61	RDRSDSPKVTGNAALTILRNCAWKKLVSKDHPKPYWDYTAIFTTRDLCGATTCDTGLM	120
DB	314	HBEQKGPEVTSNAALTILRNFCWQKHQNPSPDRDAEHYDTAIFTTRDLCGSDTCDTGLM	373
QY	121	ADYGTMCDPKRSCTVEDDGLPSAFTTAHELGHVFNPHDNVNVKVEEYFKGLRANHMSP	180
DB	374	ADVGTCDPSRSCSVIEDDGLQAAPTTAAHELGHVFNPHDDAKQACSLINGVNDSHMMAS	433
QY	181	TLIQIDRANPWSACNAIITDFLDSHGDCILLDOPSPKISIPEDLPACASYTLSSQCELAFL	240
DB	434	MUSNLDSHPWSPCSAYMITSFLDNGHECULMDKQNPQLPGDLPSTSDANRCQQTFF	493
QY	241	GVGSKPCP- YMOYCTKLMCTGKAKQWVCOTRHPFPWADGTSCEGKLCCLKAGACVHRNLN	299
DB	494	GEDSKHCPDAASTCTLMCTGTSGLVLCQTKHPFPWADGTSCEGKWCINGKCVNKTD-R	552
QY	300	KH---RVDCSWAKWDYPGSCSTCGGGVQOLARROCTNPTPANGSKYCEGVRYKRSCLNE	356
DB	553	KHDFTFPHGWSGMWGPWGDSCSTCGGGVQYTWRECDNPVPKNGKYGCEGARVTRSCNLE	612
QY	357	PCPSSASGKSFREEQCEAFNGVYHNSTNLTILAVAWPKYSGVSPDRCKLICRANGTGYF	416
DB	613	DCPDN-NGKTFREEQCEAHNEFSKASFGSGPAVEWIPKYAGVSPDKCKLICQAKGIGYF	671
QY	417	YVLAPKWVDGTLCPDSTSVQVQGCKIACGDNGLGSKKREDKCGVCGGDNKSKCKVTGL	476
DB	672	FVLQPKWVDGTPCSPDSTSVQVQGVKAGCDRIIDSKKFKDCKGCVCGGNGSTCKKISGS	731
QY	477	FTKPMHGYNFVWAIPAGASSIIDIRQYKGLIGDDNYLALANSQGYLLNGHFVVSAYER	536
DB	732	VTSAKPGYHDIITPTGATNIEVKQRNGSRNNGSFLAIKAADGTYLNGDYTLSTLEQ	791
QY	537	DLVVGSLRLYSGTGTAVESLQASRPILPLTVSVLGVGKMTTPRVRYSFYLPKPEPREDK	596
DB	792	DIMYKGVLYRISGSSALRIERSFSLPEPLTIQVLTGVNARLKPIYTFVFKK---K	847
QY	597	SSHPKDPGRGFSVLHNSVLISLNOVEQDPRDPARVWAGSWGPCSASCGSLQKRAVDCRG	656
DB	848	ES-----FNAIPTFS-----AWVIEGECSEKSELGWQRRLVECDR	884
QY	657	SAGORTVPACDAAH--RPVETQACGE-PCPTWELSAWSPSCSKCGRGFRRRSLCKVGHGG	713
DB	885	INGQ---PASECAKEVKPASTRPCADHPQCPQWLGEWSSCKTCGKGKRSKLCLSHDG	941
QY	714	RLIARDOCNLHRKPQE-LDFCVLRPC	738
DB	942	GVLSHESCDPLKKPKHFIDFCTMAEC	967

RESULT	12
AAW78189	
ID	AAW78189 standard; Protein; 967 AA.
XX	AAW78189;
XX	AC
XX	AC
DT	13-APR-1999 (first entry)
XX	XX
DE	Human secreted protein encoded by gene 64 clone HOUQC17.
XX	XX
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW	developmental abnormality; foetal deficiency; blood; allergy; renal;
KW	immune system; asthma; lymphocytic disease; brain; hepatic lymphoma;

CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. AAX04302) for increasing the stability of the fused
 CC protein as compared to the human protein only.
 CC The invention relates to 86 novel genes and their fragments (nucleic
 CC acid sequences: AAX04311-X04410; amino acid sequences AAW78126-W78225)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 86
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX04311 for described uses).

XX SQ Sequence 967 AA;
 Query Match 55.4%; Score 2241; DB 20; Length 967;
 Best Local Similarity 53.8%; Pred. No. 2e-168;
 Matches 401; Conservative 131; Mismatches 174; Indels 40; Gaps 11;
 QY 1 FVSIPIRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSTILNPIVVKVLL 60
 Db 253 FVSSHYVETMLVADQSAEFGSLKHYLLTLFSAARLXKHPXIRNSVSLVVKILVI 312
 QY 61 RDRDSGPKVTGNAALTFLNFCQAWQKLNKVDKHPYWDTAILFTRODLCGATCDTLGM 120
 Db 313 HDEQKGPVTSNAALTFLNFCQWQHPPSDRDAEHYDTAILFTRODLCGATCDTLGM 372
 QY 121 ADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNYKVEEFGKLRANHMSP 180
 Db 373 ADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNYKVEEFGKLRANHMSP 432
 QY 181 TLTIQIDRANPWSACSAIITFDLDSGHGDCILDDQPSKIPSLPEDLPAGSYTLSQOCELAF 240
 Db 433 MLNLDHSPQWSPCSAYMITSFLDNGHGBCLMDKQNPQLPGLDLPSTYDANROCOQTF 492
 QY 241 GVSKPCP-YMYCTVTKWCTKAKQOMVCQTRHFWADTSCGEGKCLKLGACVERHNLN 299
 Db 493 GEDSKPCPAASTCSPLWCTGTSGVLVCQTRHFWADTSCGEGKCLKLGACVERHNLN 356
 QY 300 KH---RVDSWAKWDPYPCSTCGGVLARRQCTNPTTPANGKYCGVRVYKBSCNLE 551
 Db 552 KHEDTFHSGWGWPGWDCSTCGGVOYTWRECDNVPKNGKYCEKRVYKBSCNLE 611
 QY 357 PCSSASGSKFREOECEAFNGYNHSTNRLTLAVAWPKYSGVSPDKCLICRANGTYG 416
 Db 612 DCPDN-NGKTFREOECEAHNEFSKASFGSGPAVEMPKYAGVSPDKCLICQAKGYG 670
 QY 417 YVLAPKVDGTLCSPDSTSVCGQKICAGCDNGLSKRFPDKCGVCGDGNKSKVIGL 476
 Db 671 FVLQPKVDGTLCSPDSTSVCGQKICAGCDRIIDSKKFKDKCGVCGDGNKSKVIGL 730
 QY 477 FTKPMHGYNFVVAIPAGASSIDIRQGYKGLIGDDNYLALKNQSKYLLNGHFVVSAYER 536
 Db 731 VTSAPGYHDIITPGATNIEVKQNRQSRNGSFLAIAKADGTIYLDGTITLSTLEQ 790
 QY 537 DLVVGSLRYSGTAVESLQASRILEPLFVLSVQKMTPPRRVYSFYLPKPREDK 596
 Db 791 DIMYKGVLYRYSGSSAALRIERSFPLKEPLTIQVLTGNALRPKIKTYFVKK- ---K 846
 QY 597 SSHPKDPRGSPVLSVLSNQVEQDDRPAPRVAGSWGPCASCGSLQKRAVDRC 656
 Db 847 ES-----FNAIFTS-----AWIEWGCSKSCELGQWRRLVECRD 883
 QY 657 SAGQRTVPACDAH--RPVETOACGE-PCPTWELSAWSPCSKSGRGFORSLKCVGHGG 713
 Db 884 INQ---PASECAKEVKPASTRPCADHPCPQWQLGEWSSCSKTCCKGYKRSKLCLSHDG 940
 QY 714 LLARQOCLNHRKPBQ-B-LDFCVLRPC 738
 Db 941 GVLSHESCDPLKKPHFIDFCTMAEC 966

RESULT 13

AA53899
 ID AAY53899 standard; Protein; 950 AA.
 AC AAY53899;
 DT 13-MAR-2000 (first entry)
 DE Amino acid sequence of a murine ADAMTS-1 protein.
 KW GON-1; metalloproteinase; cell migration; modulator; ADAMTS-1;
 KW metalloproteinase domain; thrombospondin domain; abnormal cell migration;
 KW organ shaping; sterility; cancer metastasis.
 OS Mus sp.
 PN WO9961656-A2.
 XX 02-DEC-1999.
 XX 28-MAY-1999; 99WO-US11918.
 PR 29-MAY-1998; 98US-0087170.
 PR 13-APR-1999; 99US-0129023.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 PI Kimble JE, Brelloch RH;
 XX WPI; 2000-072633/06.
 DR
 PT Identifying modulators of proteins containing metalloprotease and
 PT thrombospondin domains, potentially useful for controlling cell
 PT migration and organ shaping -
 PS Disclosure; Fig 1C; 60pp; English.
 CC The present sequence represents a murine ADAMTS-1 protein. ADAMTS-1 is
 CC a metalloproteinase. The specification describes another related
 CC metalloproteinase, a C. elegans GON-1 protein, that lacks a transmembrane
 CC domain and possesses a predicted metalloprotease domain between residues
 CC 269-436. In C. elegans hermaphrodites, GON-1 is required for migration of
 CC two distal tip cells to produce elongated tubes, whereas in males, GON-1
 CC is required for migration of a single linker cell to produce a single
 CC elongated tube. The protein is used in the method of the invention. The
 CC specification describes a method for identifying a modulator of a
 CC protein that contains a metalloproteinase domain and a thrombospondin
 CC domain. The method comprises treating a target organism, having a
 CC developing gonadal cell that is responsive to the protein, with a test
 CC compound, and determining any change in migration or shape of the cell
 CC attributable to the test compound. The compounds identified are
 CC potential therapeutic modulators of abnormal cell migration and organ
 CC shaping, e.g. for rendering animals (specifically nematodes) sterile
 CC and for inhibiting cancer metastases.
 XX Sequence 950 AA;
 SQ
 Query Match 54.9%; Score 2220; DB 21; Length 950;
 Best Local Similarity 53.3%; Pred. No. 8e-167;
 Matches 398; Conservative 130; Mismatches 177; Indels 42; Gaps 12;
 QY 1 FVSIPIRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSTILNPIVVKVLL 60
 Db 237 FVSSHYVETMLVADQSAEFGSLKHYLLTLFSAARLYRHPSTILNPIVVKVLL 296
 QY 61 RDRDSGPKVTGNAALTFLNFCQAWQKLNKVDKHPYWDTAILFTRODLCGATCDTLGM 120
 Db 297 YEEQKGPVTSNAALTFLNFCQWQHPPSDRDAEHYDTAILFTRODLCGATCDTLGM 356
 QY 121 ADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNYKVEEFGKLRANHMSP 180
 Db 357 ADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNYKVEEFGKLRANHMSP 415
 QY 181 TLTIQIDRANPWSACSAIITFDLDSGHGDCILDDQPSKIPSLPEDLPAGSYTLSQOCELAF 240

Db 416 MLSSLDHSPWSPCSAYMTSFLDNGHGECLMDKPNPILKPSDLPGTLIDANROCCFTF 475
QY 241 GVSGKPCP-YMOYCTKLWCTGKAGQWCVQTRHFPWADGTSGEGKLCGLKACVERHNLN 299
Db 476 GEESKPCDPAASTCTTLWCTGTGGVLVCTHFPWADGTSGEGKWCVCVKNKTDN- 534
QY 300 KH---RVDSWAKWDYPGCSRTCGGVQVQARRQCTNPTPANGKYCEGVRYKYSCLNLE 356
Db 535 KHFPATVHGSWGPWPGWDCSRTCGGVQVQYTMRECDNPVPKNGKYCEKRYRSCNIE 594
QY 357 PCPSSASGKSFREOCEAFNGYNHSTNRLTLAVWVPKYSVSPDRCKLCIRANGTGYF 416
Db 595 DCPDN-NGKTFREOCEAHNEFSKASFGNEPTVETPKYAGVSPKDRCKLTCEAKGIGYF 653
QY 417 YVLAPKVVDDGTLCSPDSTVCGVQKCIKAGCDNGLSKKRFDKCGVCGDNDKCKVVTGL 476
Db 654 FVLQPKVVDGTCPDSTVCGVQVQVQKAGCDRIIDSKKFKDKCGVCGDNGSTCKKMSGI 714
QY 477 FTKPMHGYNFVAIPAGASSIDIRQYKGLIGDDNYLALKNSQKYLNGHVFVSAVER 536
Db 714 VTSTRPGYHDIPTIPAGATNIEVKHNRQGRNNGSFLAIRAADGTIYILNGFTLSTLEQ 774
QY 537 DLVVKSLRYSGTGAVESLQASRPILPTVEVLSVCKMTPPRVRYSFYLPKPREDK 596
Db 774 DLYTGTVLRYSGSAALEIRSFPLKEPLTIQVLMVGHALRPKIKTYFPWKKTES-- 831
QY 597 SSHPKDPRGSPSVLHNSVLSNQVEQDDPPARWAGSWGPCASGSLQKRAVDCRG 656
Db 832 -----FNAIPTFS-----EWVIEEWGCSKTCGSGWQRRVYQCRD 867
QY 657 SAGQRTVPACDAH--RPVETOACGE-PCPTW 685
Db 867 INGH---PASECAKEVKPASTRPCADLPCHWQVGDWSPSCSKTCGRGKRCCTLKCVSHD 923
QY 713 GRLLARDQCNLHRRKQBP-LDFCVLRPC 738
Db 924 GGVLSNESCPLKPKHYIDFCHLTQC 950
RESULT 14
AAB21265
ID AAB21265 standard; Protein; 896 AA.
XX AAB21265;
AC
XX
DT 23-FEB-2001 (first entry)
XX
DE Mouse metalloproteinase ADAMTS-1.
XX
KW Mouse; ADAMTS-1; metalloproteinase; ADAM;
KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; neurotropic; neuroprotective; antiparkinsonian;
KW cerebroprotective; cytoskeletal; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.
XX
OS Mus musculus.
XX
XX WO200053774-A2.
PN
XX
PD 14-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US06237.
XX
PR 08-MAR-1999; 99US-0264585.
XX
XX (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX
XX kelner GS, Clark M, Maki RA;
XX
XX WPI; 2000-594326/56.
XX

PT Polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to
PS prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX Disclosure; Fig 17; 129pp; English.
XX
CC The present sequence is mouse metalloproteinase ADAMTS-1. The
CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
CC and Metalloproteinase Domain) family. Members of the ADAMTS family
CC contain a thrombospondin domain in addition to the disintegrin and
CC metalloproteinase domains found in the ADAMTS. ADAMTS polypeptides are
CC useful for the manufacture of medicaments for treating conditions
CC associated with neuroinflammation and/or neurodegeneration, such as
CC Alzheimer's disease, Parkinson's disease and stroke. They are also
CC useful for treating conditions associated with cell proliferation, cell
CC migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC and autoimmune diseases. They can be used to treat patients afflicted
CC with an invasive tumour, a brain tumour or brain injury.
XX
SQ Sequence 896 AA;
Query Match 51.9%; Score 2096.5; DB 21; Length 896;
Best Local Similarity 54.2%; Pred. No. 4.8e-157;
Matches 375; Conservative 115; Mismatches 163; Indels 39; Gaps 9;
QY 1 FVSIPIRYVETLVVADSNVKEHGDLEHYLLTLATARLHPSILNPIVIVVVKVLL 60
Db 237 FVSSPIRYVETMLVADSNVKEHGDLEHYLLTLATARLHPSILNPIVIVVVKVLL 296
QY 61 RDRSDSPKVTGNAALTNRNFCWQKKNVSKHPEYWDATAILFTRODLCGATCTDGLM 120
Db 297 YEEQKPEVTSNAALTNRNFCWQKKNVSKHPEYWDATAILFTRODLCGATCTDGLM 356
QY 121 ADVGTMCDPKRSCSVIEDGLPSAFTTAHELGHVFNMPHNVKVEEFGKLRANHMSP 180
Db 357 ADVGTMCDPKRSCSVIEDGLPSAFTTAHELGHVFNMPHNVKVEEFGKLRANHMSP 416
QY 181 TLIQIDRANPWSACSAIITDLDSDHGDCLLDOPSKPISLPEDLPAGASYTLSSQCELA 240
Db 417 MLSSLDHSPWSPCSAYMTSFLDNGHGECLMDKPNPILKPSDLPGTLIDANROCCFTF 476
QY 241 GVSGKPCP-YMOYCTKLWCTGKAGQWCVQTRHFPWADGTSGEGKLCGLKACVERHNLN 299
Db 477 GEESKPCDPAASTCTTLWCTGTGGVLVCTHFPWADGTSGEGKWCVCVKNKTDN- 535
QY 300 KH---RVDSWAKWDYPGCSRTCGGVQVQARRQCTNPTPANGKYCEGVRYKYSCLNLE 356
Db 535 KHFPATVHGSWGPWPGWDCSRTCGGVQVQYTMRECDNPVPKNGKYCEKRYRSCNIE 595
QY 357 PCPSSASGKSFREOCEAFNGYNHSTNRLTLAVWVPKYSVSPDRCKLCIRANGTGYF 416
Db 595 DCPDN-NGKTFREOCEAHNEFSKASFGNEPTVETPKYAGVSPKDRCKLTCEAKGIGYF 654
QY 417 YVLAPKVVDDGTLCSPDSTVCGVQKCIKAGCDNGLSKKRFDKCGVCGDNDKCKVVTGL 476
Db 654 FVLQPKVVDGTCPDSTVCGVQVQVQKAGCDRIIDSKKFKDKCGVCGDNGSTCKKMSGI 714
QY 477 FTKPMHGYNFVAIPAGASSIDIRQYKGLIGDDNYLALKNSQKYLNGHVFVSAVER 536
Db 714 VTSTRPGYHDIPTIPAGATNIEVKHNRQGRNNGSFLAIRAADGTIYILNGFTLSTLEQ 774
QY 537 DLVVKSLRYSGTGAVESLQASRPILPTVEVLSVCKMTPPRVRYSFYLPKPREDK 596
Db 774 DLYTGTVLRYSGSAALEIRSFPLKEPLTIQVLMVGHALRPKIKTYFPWKKTES-- 832
QY 597 SSHPKDPRGSPSVLHNSVLSNQVEQDDPPARWAGSWGPCASGSLQKRAVDCRG 656
Db 832 -----FNAIPTFS-----EWVIEEWGCSKTCGSGWQRRVYQCRD 867
QY 657 SAGQRTVPACDAH--RPVETOACGE-PCPTW 685
Db 867 INGH---PASECAKEVKPASTRPCADLPCHW 896


```

RESULT 15
AAB21261
ID AAB21261 standard; Protein; 381 AA.
XX
AC AAB21261;
XX
DT 23-FEB-2001 (first entry)
XX
DE Human metalloproteinase ADAMTS-5.
XX
KW Human; ADAMTS-5; metalloproteinase; ADAM;
KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; neurotrophic; neuroprotective; antiparkinsonian;
KW cerebroprotective; cytostatic; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.
XX
OS Homo sapiens.
XX
FT Key Location/Qualifiers
FT Misc-difference 196 /note= "Xaa= any amino acid"
FT Misc-difference 268 /note= "Xaa= any amino acid"
FT Misc-difference 304 /note= "Xaa= any amino acid"
FT Misc-difference 308 /note= "Xaa= any amino acid"
FT Misc-difference 308 /note= "Xaa= any amino acid"
XX
WO200053774-A2.
XX
PN 14-SEP-2000.
XX
PD 08-MAR-2000; 2000WO-US06237.
XX
PF 08-MAR-1999; 99US-0264585.
XX
PR (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX
PA Kelner GS, Clark M, Maki RA;
XX
PI WPI; 2000-594326/56.
XX
DR N-PSDB; AAA95831.
XX
PT Polynucleotide encoding novel members of a disintegrin,
XX metalloproteinase and thrombospondin domain protein family used to
XX prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX Claim 12; Fig 23; 129pp; English.
XX
CC The present sequence is human metalloproteinase ADAMTS-5. The
XX ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
XX and Metalloproteinase Domain) family. Members of the ADAMTS family
XX contain a thrombospondin domain in addition to the disintegrin and
XX metalloproteinase domains found in the ADAMTS. ADAMTS polypeptides are
XX useful for the manufacture of medicaments for treating conditions
XX associated with neuroinflammation and/or neurodegeneration, such as
XX Alzheimer's disease, Parkinson's disease and stroke. They are also
XX useful for treating conditions associated with cell proliferation, cell
XX migration, inflammation and/or angiogenesis, such as cancer, arthritis
XX and autoimmune diseases. They can be used to treat patients afflicted
XX with an invasive tumour, a brain tumour or brain injury.
XX
SQ Sequence 381 AA;

```

```

Query Match 48.6%; Score 1966; DB 21; Length 381;
Best Local Similarity 94.0%; Pred. No. 3.3e-147;
Matches 358; Conservative 3; Mismatches 18; Indels 2; Gaps 2;
XX
QY 28 HYLTLTLLTAARLYRHPSTILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWQKLL 87
||| |:::|||||
DB 2 HYRARAARAG-IFKHPSILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWQKLL 60

```

Search completed: April 29, 2003, 17:20:46
Job time : 45.9135 secs

```

QY 88 NKVSDKHPEYWDTAILFTRODLGCGATTCTDLGMADVGTMCDPKRSCSVIEDDGLPSAFTT 147
||| |:::|||||
DB 61 NKVSDKHPEYWDTAILFTRODLGCGATTCTDLGMADVGTMCDPKRSCSVIEDDGLPSAFTT 120
||| |:::|||||
QY 148 AHGLGHVFNPHDNVNVKCEEVFGKLRANHMSPSTLIQIDRANPWSACSAAIITDFLDLSDGH 207
||| |:::|||||
DB 121 AHGLGHVFNPHDNVNVKCEEVFGKLRANHMSPSTLIQIDRANPWSACSAAIITDFLDLSDGH 180
||| |:::|||||
QY 208 GDCLLDQPSKPISLPEDLPGASYTLSSQCELAFGVSGSKPCPYMOCYCTKWLCTCKAKGMV 267
||| |:::|||||
DB 181 GDCLLDQPSKPIFLPDLPGASYTLSSQCELAFGVSGSKPCPYMOCYCTKWLCTCKAKGMV 240
||| |:::|||||
QY 268 CQTRHFFWADGTSCEGKGLCKLKGACVERHNLNKHVRVDGSKWAKWDYPGCSRTCGGGVQLA 327
||| |:::|||||
DB 241 CQTRHFFWADGTSCEGKGLCKLKGACVEXHNLNKHVRVDGSKWAKWDYPGCSRTCGGGVQLA 300
||| |:::|||||
QY 328 RROCTNPTP-ANGKCYCEGVVRVYRSCNLEPCPSSASGKSFREEOCEAFNGYNHSTNRLT 386
||| |:::|||||
DB 301 RROXHQPLPTGGKYCEGVVRVYRSCNLEPCPSSASGKSFREEOCEAFNGYNHSTNRLT 360
||| |:::|||||
QY 387 LAVAVVPKYSVSPRDKCKLI 407
||| |:::|||||
DB 361 LAVAVVPKYSVSPRDKCKLI 381

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:25 ; Search time 46.2301 Seconds
(without alignments)
2738.218 Million cell updates/sec

Title: US-10-009-332-1
Perfect score: 5164
Sequence: 1 MLLGLITLAFAGRTAGGFE.....DQCNLHRKPEQLDFCVLRPC 950

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5164	100.0	950	22	AAG62299 Human metalloprotease #2.
2	5156	99.8	950	23	AAE22541 Human metalloprotease PRT
3	5111	99.0	952	23	AAU74751 Human metalloprotease
4	4856.5	94.0	928	23	AAU72899 Rat metalloprotease
5	2497.5	48.4	505	21	AAE21257 Human METH1 protei
6	2480.5	48.0	950	20	AAU49501 Human ADAM-type me
7	2480.5	48.0	950	22	AAU73549 Human METH1. Homo
8	2480.5	48.0	950	22	AAU50002 Human integrin lig
9	2480.5	48.0	967	19	AAW80285 Protein; SEQ ID 12
10	2480.5	48.0	968	22	AAU50011

11	2479.5	48.0	967	20	AAU04142 Human Tango-71 pro
12	2458.5	47.6	967	20	AAU78189 Human secreted pro
13	2445.5	47.4	950	21	AAU53899 Amino acid sequenc
14	2322	45.0	896	21	AAE21265 Mouse metalloprote
15	2274	44.0	727	20	AAU78435 Human ADAMTS-1 pro
16	2135.5	41.4	890	20	AAU49502 Human METH2 protei
17	2135.5	41.4	890	22	AAU50003 Human METH2. Homo
18	2126.5	41.2	889	22	AAU74946 Murine ADAM type me
19	2113	40.9	905	22	AAU72284 Murine ADAMTS-8 am
20	1966	38.1	381	21	AAE21261 Human metalloprote
21	1965	38.1	1629	23	ABG30703 Human aggrecanase
22	1965	38.1	1629	23	AAO14448 Human ADAMTS-SI pr
23	1965	38.1	1916	23	AAU19173 Human metalloprote
24	1965	38.1	1935	23	AAU72896 Human aggrecanase
25	1957	37.9	1602	23	ABG30702 Human metalloprote
26	1943.5	37.6	1073	21	AAE21264 Human ADAMTS-9 alt
27	1939.5	37.6	1934	22	AAU72301 Human aggrecan deg
28	1917	37.1	837	22	AAU75425 Human aggrecan
29	1916	37.1	837	21	AAU99429 Human PRO1563 (UNQ
30	1916	37.1	837	22	AAU29199 Protein of the inv
31	1916	37.1	837	22	AAU66178 Human aggrecanase-
32	1915	37.1	837	22	AAU78228 Human metalloprote
33	1915	37.1	840	21	AAE21256 Human metalloprote
34	1913	37.0	367	23	AAE22542 Human metalloprote
35	1903	36.9	680	21	AAE21251 Human metalloprote
36	1882	36.4	1907	23	AAU77133 Human metalloprote
37	1874	36.3	1505	23	AAU72897 Human metalloprote
38	1872	36.3	947	22	AAU86950 Murine ADAMTS-5 am
39	1854	35.9	930	22	AAU72280 Human ADAMTS-9 am
40	1848.5	35.8	1882	22	AAU72286 Human aggrecan deg
41	1847.5	35.8	930	20	AAU75426 Human aggrecan
42	1840	35.6	929	21	AAU41226 Human ORFX ORF990
43	1798.5	34.8	874	22	AAU72287 Murine ADAMTS-9 am
44	1783	34.5	958	21	AAE21255 Human metalloprote
45	1762	34.1	870	21	AAE21252 Rat metalloprote

ALIGNMENTS

RESULT 1
AAG62299
ID AAG62299 standard; protein; 950 AA.
XX AC
XX AAG62299;
XX AC
XX DT 23-AUG-2001 (first entry)
XX DT
XX Human metalloprotease MDT56 protein.
DE Human metalloprotease; human; aggrecanase; joint disease; osteoarthritis;
KW osteopathic; antiarthritic.
XX KW
XX Homo sapiens.
XX OS
XX WO200134785-A1.
XX PN
XX 17-MAY-2001.
XX PD
XX 10-NOV-2000; 2000WO-JP07917.
XX PF
XX 11-NOV-1999; 99JP-0321740.
XX PR 16-MAY-2000; 2000JP-0144020.
XX PR
XX (YAMA) YAMANOUCHI PHARM CO LTD.
XX (KAZU-) KAZUSA DNA RES INST.
XX PA
XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
XX WPI; 2001-343602/36.
XX DR N-PSDB; AAH41003.
XX DR
XX Metalloprotease with aggrecanase activity for treating joint diseases
XX PT

PT especially osteoarthritis -
 PS Claim 1; Page 56-60; 85pp; Japanese.
 XX This invention relates to a metalloprotease with aggreganase activity.
 CC The invention includes protein and DNA sequences of the metalloprotease,
 CC vectors containing the DNA, host cells transformed by the vectors, and
 CC antibodies directed against the metalloprotease. The antibodies, protein
 CC and DNA sequences can be used in the treatment and prevention of joint
 CC diseases, particularly osteoarthritis. The treatment may result in
 CC osteopathic and antiarthritic activity. The present sequence represents
 CC the metalloprotease of the invention termed MDT56.
 XX
 SQ Sequence 950 AA;
 Query Match 100.0%; Score 5164; DB 22; Length 950;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 950; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLLGILTLAFAGTAGGFEPEVWVPIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
 DB 1 MLLGILTLAFAGTAGGFEPEVWVPIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
 QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFCYSGDVNAEPDFAAVSLC 120
 DB 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFCYSGDVNAEPDFAAVSLC 120
 QY 121 GGLRGAFYGAEEYVISPPLNAPAAQNSQGAHLQRRGVPDSDPTSRGCVASGW 180
 DB 121 GGLRGAFYGAEEYVISPPLNAPAAQNSQGAHLQRRGVPDSDPTSRGCVASGW 180
 QY 181 NPAILRALDPKPRAGFESRRSRGRKRFRFVPIRYVETLVVADESVMVKEHGALEH 240
 DB 181 NPAILRALDPKPRAGFESRRSRGRKRFRFVPIRYVETLVVADESVMVKEHGALEH 240
 QY 241 YLLTLATAARLYRHPSTILNINIVVKVLLRRDSDGPKVTGNAALTFRNFCWQKLN 300
 DB 241 YLLTLATAARLYRHPSTILNINIVVKVLLRRDSDGPKVTGNAALTFRNFCWQKLN 300
 QY 301 KYSDKHPEYWDATILFTQDLCGATCTDGLMADVGTMCDPKRSCVIEDDGLPSAFTTA 360
 DB 301 KYSDKHPEYWDATILFTQDLCGATCTDGLMADVGTMCDPKRSCVIEDDGLPSAFTTA 360
 QY 361 HELGHVFNPHNDNVKCEVEFGKLRANHMSPTLIQIDRANPWSACSAIIITDFLDSGHG 420
 DB 361 HELGHVFNPHNDNVKCEVEFGKLRANHMSPTLIQIDRANPWSACSAIIITDFLDSGHG 420
 QY 421 DCLLDQPSKPTSLPDLPGASYTLISQCELAFLGVGSKPCPYMQYCTKLWCTGKAGQWVC 480
 DB 421 DCLLDQPSKPTSLPDLPGASYTLISQCELAFLGVGSKPCPYMQYCTKLWCTGKAGQWVC 480
 QY 481 QTRHFPWADGTSCGSGKLCCLGACVERHNKLRHVRDGSWAKWDPYGPCSRTCGGVQLAR 540
 DB 481 QTRHFPWADGTSCGSGKLCCLGACVERHNKLRHVRDGSWAKWDPYGPCSRTCGGVQLAR 540
 QY 541 RQCTNPTPANGKYCEGVVRYRSCNLEPCPSASGKSFREOCEAFNGYHNSTNRLTLA 600
 DB 541 RQCTNPTPANGKYCEGVVRYRSCNLEPCPSASGKSFREOCEAFNGYHNSTNRLTLA 600
 QY 601 VAWVPKYSVSPDRCKKLCIRANGTYFYVLAPKVVDGTLGSPDSTVCVQKCIKAGCD 660
 DB 601 VAWVPKYSVSPDRCKKLCIRANGTYFYVLAPKVVDGTLGSPDSTVCVQKCIKAGCD 660
 QY 661 GNLSKKRPDKCGVCGGDNKSKVYGLFTKPMHGYNFVVAIPAGASSIDIRQRYKGLI 720
 DB 661 GNLSKKRPDKCGVCGGDNKSKVYGLFTKPMHGYNFVVAIPAGASSIDIRQRYKGLI 720
 QY 721 GDDNYLALNKSOGKYLNLNGHFVSAVERDLVVKGSLRYSGTGAVESLQASRPILPLET 780
 DB 721 GDDNYLALNKSOGKYLNLNGHFVSAVERDLVVKGSLRYSGTGAVESLQASRPILPLET 780
 QY 781 VEVLSVGMKTPPRVRYSFYLPKEPREDKSSHPKDPGRPSVLHNSVLSLSNOVEQPDPRPP 840
 DB 781 VEVLSVGMKTPPRVRYSFYLPKEPREDKSSHPKDPGRPSVLHNSVLSLSNOVEQPDPRPP 840

DB 781 VEVLSVGMKTPPRVRYSFYLPKEPREDKSSHPKDPGRPSVLHNSVLSLSNOVEQPDPRPP 840
 QY 841 ARWVAGSWGSPCSASCSGSLQKRAVDCRGSAQORTVPACDAHRPVETQACGEPCTWELS 900
 DB 841 ARWVAGSWGSPCSASCSGSLQKRAVDCRGSAQORTVPACDAHRPVETQACGEPCTWELS 900
 QY 901 ANSPCKSGCRGFORSLKCVGHGGRLLARDQCNLHRKPQELDFCVLRPC 950
 DB 901 ANSPCKSGCRGFORSLKCVGHGGRLLARDQCNLHRKPQELDFCVLRPC 950
 RESULT 2
 AAE22541
 ID AAE22541 standard; Protein; 950 AA.
 XX
 AC AAE22541;
 XX
 DT 26-JUL-2002 (first entry)
 XX
 DE Human protease #2.
 KW Human; novel human protein; NHP; protease; biological disorder; obesity;
 KW high blood pressure; arthritis; connective tissue disorder; infertility;
 KW gene therapy; enzyme.
 XX
 OS Homo sapiens.
 XX
 PN WO200226949-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 27-SEP-2001; 2001WO-US30350.
 XX
 PR 29-SEP-2000; 2000US-236689P.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Friddle CJ, Hilbun E;
 DR WPI; 2002-372123/40.
 DR N-PSDB; AAD35569.
 XX
 PT Novel nucleic acid encoding a human protease, useful as a hybridization
 PT probe for screening libraries and assessing gene expression patterns -
 XX
 PS Claim 6; Page 36-38; 4lpp; English.
 CC
 CC The present sequence is novel human protein (NHP), human protease.
 CC NHPs share structural similarity with animal proteases particularly
 CC zinc metalloproteases. Sequences of the invention are useful in
 CC therapeutic, diagnostic and pharmacogenomic applications. NHP
 CC polynucleotides are used as hybridisation probes for screening
 CC libraries and assessing gene expression patterns. They can also be
 CC used for treating related biological disorders such as obesity, high
 CC blood pressure, arthritis, connective tissue disorders and infertility.
 CC They are also used in gene therapy.
 XX
 SQ Sequence 950 AA;
 Query Match 99.8%; Score 5156; DB 23; Length 950;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 949; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MLLGILTLAFAGTAGGFEPEVWVPIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
 DB 1 MLLGILTLAFAGTAGGFEPEVWVPIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
 QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFCYSGDVNAEPDFAAVSLC 120
 DB 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFCYSGDVNAEPDFAAVSLC 120
 QY 121 GGLRGAFYGAEEYVISPPLNAPAAQNSQGAHLQRRGVPDSDPTSRGCVASGW 180
 DB 121 GGLRGAFYGAEEYVISPPLNAPAAQNSQGAHLQRRGVPDSDPTSRGCVASGW 180

Dd	121	GGLRGAFGYRGAEYVISPLPNASAPAAQRNSQGAHLQRRGVPGSPGDPTSRCGVASGW	180
Qy	181	NPAILRALDYPKPRAGGEGESRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEH	240
Dd	181	NPAILRALDYPKPRAGGEGESRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEH	240
Qy	241	YLLTLATAARLYRHPSILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWOKKLN	300
Dd	241	YLLTLATAARLYRHPSILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWOKKLN	300
Qy	301	KVSDKHPEYWDTAIFTRQDLGGATCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTA	360
Dd	301	KVSDKHPEYWDTAIFTRQDLGGATCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTA	360
Qy	361	HELGHVFNNPHDNVAVCEVFGKLRANHMMSPTLIQIDRANPWSACSAAIITDFLDSHG	420
Dd	361	HELGHVFNNPHDNVAVCEVFGKLRANHMMSPTLIQIDRANPWSACSAAIITDFLDSHG	420
Qy	421	DCLLDQPSKPISLPEDLPASYYTLSSQCELAFAFGVSGKPCPYMQYCTKLWCTGKAKQWVC	480
Dd	421	DCLLDQPSKPISLPEDLPASYYTLSSQCELAFAFGVSGKPCPYMQYCTKLWCTGKAKQWVC	480
Qy	481	QTRHFPWADGTSCEGSEKCLKACACVERHNLNKHHRVDGSAKWDPYGPCSRCTCGGVQLAR	540
Dd	481	QTRHFPWADGTSCEGSEKCLKACACVERHNLNKHHRVDGSAKWDPYGPCSRCTCGGVQLAR	540
Qy	541	ROCTNPTPANGKGYCEGVVKYRSCNLEPCPSSASGKSFREOCEAFNGYNHSTNRLTLA	600
Dd	541	ROCTNPTPANGKGYCEGVVKYRSCNLEPCPSSASGKSFREOCEAFNGYNHSTNRLTLA	600
Qy	601	VAVPKYSGVSPDRCKKLCIRANGTYFYVLAPKVVDGTLGSPDSTSVCGQKCIKAGCD	660
Dd	601	VAVPKYSGVSPDRCKKLCIRANGTYFYVLAPKVVDGTLGSPDSTSVCGQKCIKAGCD	660
Qy	661	GNLSGKRRDKCGVGGDNKCKYTGFTKPMHGYNFVVAIPAGASSIDIRQGYKGLI	720
Dd	661	GNLSGKRRDKCGVGGDNKCKYTGFTKPMHGYNFVVAIPAGASSIDIRQGYKGLI	720
Qy	721	GDNYLALNSOGKYLNGHGFVYSVERDLVVKGLSRYSGTGTAVESLQASRPILPLEPT	780
Dd	721	GDNYLALNSOGKYLNGHGFVYSVERDLVVKGLSRYSGTGTAVESLQASRPILPLEPT	780
Qy	781	VEVLSGKMTTPRVRYSYFLPKPEPREDKSHPKDPRGVSVLHNSVLSNQVEQDDRRPP	840
Dd	781	VEVLSGKMTTPRVRYSYFLPKPEPREDKSHPKDPRGVSVLHNSVLSNQVEQDDRRPP	840
Qy	841	ARWVAGSWGPCASGSGGLQKRAVDCRSAGORTVPACDAHRPVETOACGECPTWELS	900
Dd	841	ARWVAGSWGPCASGSGGLQKRAVDCRSAGORTVPACDAHRPVETOACGECPTWELS	900
Qy	901	ANSPCSKSGRGFRRLSKCVGHGGRLLARDQCNLHRRKPQELDFCVLRPC	950
Dd	901	ANSPCSKSGRGFRRLSKCVGHGGRLLARDQCNLHRRKPQELDFCVLRPC	950
RESULT 3			
ID	AAU74751		
XX-	AAU74751 standard; Protein; 952 AA.		
AC	AAU74751;		
XX			
DT	09-APR-2002 (first entry)		
XX			
DE	Human protease PRTS-11 protein sequence.		
KW	Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;		
KW	cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;		
KW	inflammatory disorder; acquired immunodeficiency syndrome; AIDS;		
KW	cell proliferative disorder; developmental disorder; epilepsy;		
KW	Duchenne muscular dystrophy; epithelial disorder; neurological disorder;		
KW	reproductive disorder; endometriosis.		
OS	Homo sapiens.		

XX	WO200198468-A2.		
PN			
XX			
PD	27-DEC-2001.		
XX			
PF	13-JUN-2001; 2001WO-US19178.		
XX			
PR	16-JUN-2000; 2000US-212336P.		
PR	22-JUN-2000; 2000US-213955P.		
PR	29-JUN-2000; 2000US-215396P.		
PR	07-JUL-2000; 2000US-216821P.		
PR	14-JUL-2000; 2000US-218946P.		
XX	(INCY-) INCYTE GENOMICS INC.		
PA			
XX	Yue H, Elliott VS, Gandhi AR, Lal P, Au-Young J, Tribouley CM;		
PI	Deleagean AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;		
PI	Walia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;		
PI	Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;		
PI	Kallick DA;		
XX			
DR	WPI; 2002-090437/12.		
DR	N-PSDB; ABK12894.		
XX			
PT	Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful		
PT	in the diagnosis, treatment and prevention of gastrointestinal (e.g.		
PT	gastritis), cardiovascular (e.g. atherosclerosis) and cell		
PT	proliferative (e.g. cancer) disorders -		
XX			
PS	Claim 1; Page 144-146; 177pp; English.		
XX			
CC	The present invention relates to twenty one new human proteases,		
CC	referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and		
CC	polypeptides of the invention are useful in the diagnosis, treatment and		
CC	prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and		
CC	Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and		
CC	myocardial infarction, autoimmune/inflammatory e.g. acquired		
CC	immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell		
CC	proliferative e.g. cancer, developmental e.g. Duchenne and Becker		
CC	muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.		
CC	epilepsy and Alzheimer's disease and reproductive e.g. infertility and		
CC	endometriosis disorders. Numerous other examples of each disorder are		
CC	given in the specification. The present protein sequence represents		
CC	the human protease PRTS-11 protein of the invention.		
XX			
SQ	Sequence 952 AA;		
Query Match	99.0%; Score 5111; DB 23; Length 952;		
Best Local Similarity	99.4%; Pred. No. 0;		
Matches	946; Conservative 0; Mismatches 4; Indels 2; Gaps 2;		
Qy	1 MLLGLITLAFAGTAGGEPEREVVVPIRLDPDINGRRYWRGPDSDGGLIFQITAF	60	
Dd	1 MLLGLITLAFAGTAGGEPEREVVVPIRLDPDINGRRYWRGPDSDGGLIFQITAF	60	
Qy	61 QEDFYHLTPDAQFLAPAFSTEHLGVLPQGLTGSSDLRRCFYSGDVNAEPDFAVSLC	120	
Dd	61 QEDFYHLTPDAQFLAPAFSTEHLGVLPQGLTGSSDLRRCFYSGDVNAEPDFAVSLC	120	
Qy	121 GGLRGAFGYRGAEYVISPLPNASAPAAQRNSQGAHLQRRGVPGSPGDPTSRCGVASGW	180	
Dd	121 GGLRGAFGYRGAEYVISPLPNASAPAAQRNSQGAHLQRRGVPGSPGDPTSRCGVASGW	180	
Qy	181 NPAILRALDYPKPRAGGEGESRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEH	240	
Dd	181 NPAILRALDYPKPRAGGEGESRRSRGRKRFVSIPIRYVETLVVADESVMKFGADLEH	240	
Qy	241 YLLTLATAARLYRHPSILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWOKKLN	300	
Dd	241 YLLTLATAARLYRHPSILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWOKKLN	300	
Qy	301 KVS DKHPEYWDTAIFTRQDLGGATCDTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTA	360	

Db 301 KVSQKHEPYWDTAILFTRODLGATTCDTLGMADVMTCDPKRSCSVIEDDGLPSAFTTA 360
 QY 361 HELGHVFNMPHDNVKCEEEVFGKLRANHMSPPTLIQIDRANPWSACSAAITDLDLSHG 420
 Db 361 HELGHVFNMPHDNVKCEEEVFGKLRANHMSPPTLIQIDRANPWSACSAAITDLDLSHG 420
 QY 421 DCLLDQSPKPSLPELPGASYTLSQCELAFCGVGSKPCPYMOYCTKLMCTGKAKQWVC 480
 Db 421 DCLLDQSPKPSLPELPGASYTLSQCELAFCGVGSKPCPYMOYCTKLMCTGKAKQWVC 480
 QY 481 QTRHFPWADTSCGEGKLCIKGACVERHNLNKRVDSWAKWDPYPCSRCTGGGVQLAR 540
 Db 481 QTRHFPWADTSCGEGKLCIKGACVERHNLNKRVDSWAKWDPYPCSRCTGGGVQLAR 540
 QY 541 RQCTNPTPANGKYCEGVRYKYSCLNLEPCPSASCKSPREOCEAFNGYNHSTNRLTLA 600
 Db 541 RQCTNPTPANGKYCEGVRYKYSCLNLEPCPSASCKSPREOCEAFNGYNHSTNRLTLA 600
 QY 601 VAWPKYSGVSPRDKCLICRANGTYFYVYLAPK - VVDGTLCPDSTSVCVQKCIKAGC 659
 Db 601 VAWPKYSGVSPRDKCLICRANGTYFYVYLAPKVVVDGTLCPDSTSVCVQKCIKAGC 660
 QY 660 DGNLGSKKRDKCGVCGGDNKCKVTGLFTKPMHGYNFVVAIPAGASSIDIRQYKGL 719
 Db 661 DGNLGSKKRDKCGVCGGDNKCKVTGLFTKPMHGYNFVVAIPAGASSIDIRQYKGL 720
 QY 720 IGDNYLALANSQKYLNLNGHFVVSVERDLVVKGSLLRYSGTGAVESLQASRPILPEL 779
 Db 721 IGDNYLALANSQKYLNLNGHFVVSVERDLVVKGSLLRYSGTGAVESLQASRPILPEL 780
 QY 780 TVEVLSVGKMTPPRVRYSFYLPKEPREDKSSHPKDPR - GPSVLHNSVLSNQVEOPDDR 838
 Db 781 TVEVLSVGKMTPPRVRYSFYLPKEPREDKSSHPPHPPGPGPSVLHNSVLSNQVEOPDDR 840
 QY 839 PPARWAGSGPCASGSGGLQRAVDCRSGAGQRTVPACDAHRPVEVQACGECPTWE 898
 Db 841 PPARWAGSGPCASGSGGLQRAVDCRSGAGQRTVPACDAHRPVEVQACGECPTWE 900
 QY 899 LSAPSPCKSCGRGFORRLSKYGHGRLRLARDQCNLHRKPQLDPCVLRPC 950
 Db 901 LSAPSPCKSCGRGFORRLSKYGHGRLRLARDQCNLHRKPQLDPCVLRPC 952

RESULT 4
 AAU72899
 ID AAU72899 standard; Protein: 928 AA.
 AC AAU72899;
 AC AAU72899;
 DT 26-FEB-2002 (first entry)
 XX Human metalloprotease partial protein sequence #11.
 DE Human; protease; PCR primer; cytosolic; immunomodulator; cardiac;
 KW vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;
 KW hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;
 KW anorectic; antinflammatory; aspartyl protease; cysteine protease;
 KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;
 KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;
 KW immune-related disease; cardiovascular disease; neuronal disease;
 KW migraine; sexual dysfunction; mood disorder; attention disorder;
 KW cognition disorder; hypotension; hypertension; psychotic disorder;
 KW dyskinesia; metabolic disorder; inflammatory disorder.
 XX Homo sapiens.
 OS
 PN W0200183782-A2.
 XX
 PD 08-NOV-2001.
 XX
 PF 04-MAY-2001; 2001WO-US14431.
 XX
 PR 04-MAY-2000; 2000US-201879P.

(SUGE-) SUGEN INC.
 Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
 Payne V;
 WPI; 2002-041502/05.
 DR N-PSDB; AAS97182.
 XX Novel protease polypeptide useful for screening for substances that may
 be used to treat, e.g., cancers, immune-related diseases,
 PT cardiovascular disease, migraine, pain, psychotic and inflammatory
 PT disorders -
 XX Claim 28; Figure 2G; 232pp; English.
 XX The invention relates to an isolated, enriched, or purified protease
 CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
 CC screen for substances (S) that may modulate its activity. Administering
 CC S (which modulates protease activity in vitro) may be used to treat a
 CC disease or disorder selected from cancers (e.g., of tissues, of blood or
 CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,
 CC brain, ovarian, bladder or kidney), immune-related diseases and
 CC disorders, cardiovascular disease, brain or neuronal-associated diseases
 CC (e.g., central or peripheral nervous system diseases, migraine, pain,
 CC sexual dysfunction, mood disorders, attention disorders, cognition
 CC disorders, hypotension, hypertension, psychotic disorders, neurological
 CC disorders and dyskinesias), metabolic disorders and inflammatory
 CC disorders. (I) may also be useful as a diagnostic tool for a disease or
 CC disorder such as those above. AAU72876-AAU72910 represent human
 CC protease amino acid sequences of the invention.
 XX
 SQ Sequence 928 AA;
 Query Match 94.0%; Score 4856.5; DB 23; Length 928;
 Best Local Similarity 94.7%; Pred. No. 0;
 Matches 901; Conservative 0; Mismatches 1; Indels 49; Gaps 2;
 QY 1 MLLLGILTLAFAGTAGGFEPEYVVPVIRLDPDINGRRYWRGPDGGLQIFQITAF 60
 Db 26 MLLLGILTLAFAGTAGGFEPEYVVPVIRLDPDINGRRYWRGPDGGLQIFQITAF 85
 QY 61 QEDFYHLHTPDQAFLAPAFSTEHLGVPLQGLTGSSDLRRCFYSGDVNAEPDFAAAYSLC 120
 Db 86 QEDFYHLHTPDQAFLAPAFSTEHLGVPLQGLTGSSDLRRCFYSGDVNAEPDFAAAYSLC 145
 QY 121 GGLGAFGYGAEYVISPLPNASAPAAQNSQGAHLQLRQGVPGSPGSDPTSRGCVASGW 180
 Db 146 GGLGAFGYGAEYVISPLPNASAPAAQNSQGAHLQLRQGVPGSPGSDPTSRGCVASGW 205
 QY 181 NPAILRALDPYKRRAGGESSRRRRSGRAKRFVSIPIRYVETLVVADESVMKFGADLEH 240
 Db 206 NPAILRALDPYKRRAGGESSRRRRSGRAKRFVSIPIRYVETLVVADESVMKFGADLEH 265
 QY 241 YLLTLTAARLYRHPSTILNINIVVXVLLLRDSDGPKVTGNALTLRNFCAWQKLN 300
 Db 266 YLLTLTAARLYRHPSTILNINIVVXVLLLRDSDGPKVTGNALTLRNFCAWQKLN 325
 QY 301 KVSQKHEPYWDTAILFTRODLGATTCDTLGMADVMTCDPKRSCSVIEDDGLPSAFTTA 360
 Db 326 KVSQKHEPYWDTAILFTRODLGATTCDTLGMADVMTCDPKRSCSVIEDDGLPSAFTTA 385
 QY 361 HELGHVFNMPHDNVKCEEEVFGKLRANHMSPPTLIQIDRANPWSACSAAITDLDLSHG 420
 Db 386 HELGHVFNMPHDNVKCEEEVFGKLRANHMSPPTLIQIDRANPWSACSAAITDLDLSHG 445
 QY 421 DCLLDQSPKPSLPELPGASYTLSQCELAFCGVGSKPCPYMOYCTKLMCTGKAKQWVC 480
 Db 446 DCLLDQSPKPSLPELPGASYTLSQCELAFCGVGSKPCPYMOYCTKLMCTGKAKQWVC 505
 QY 481 QTRHFPWADTSCGEGKLCIKGACVERHNLNKRVDSWAKWDPYPCSRCTGGGVQLAR 540
 Db 506 QTRHFPWADTSCGEGKLCIKGACVERHNLNKRVDSWAKWDPYPCSRCTGGGVQLAR 565

Db	351	DTLGMADVTGCDPSRSCSVETDDGLQAAFTTAHELGHVFNPHPHDDAKQACASLNGVND05	410
Qy	388	HMMSTLLQIDRANPWSACSAAIITDFLDGSHGDCLLDOPSKPISLPEDLPQASVYLSQQ	447
Db	411	HMAWMLNLDHSOPWSPCSAYMITTSLDNGHGECLMDKPQNPQLPGDLPGTSDANRQ	470
Qy	448	CELAFEGGSKPCP-YMQVCTTLCWCTKAKAQMYCQTRHFFPWADGTSCEGKGLCLKAGACVE	506
Db	471	CQFTFGEDSKHCPDAASTCTLWCTGTSGGVLVCQTKHFPWADGTSCEGSKWCKINGKCVN	530
Qy	507	RINLNKHK---RVDGSAWKADYPGPCSTRTCGGVQLARRQCTNPTPANGKGYCEGVVRKYR	563
Db	531	KTD-RKHEDTTPHGGWGMWGPWCSTRTCGGVQYTMRECDNPVKNKGKGYCEGRKRYR	589
Qy	564	SCNLEPCFSSAGSKFREEOCEAFNGYNHSTNRLTLAVAWPKYSGVSPDRCKCLICRAN	623
Db	590	SCNLEDCPN-NGKTFREOCEAHNEFSKASFGSGPAVEWIPKYAGVSPKDRCKLICQAK	648
Qy	624	GTGYFYVLAPKVVDTLCSPOSTSVYCVQKCIIRAGCDNGLGSKKRFDKGCVCGDNKSK	683
Db	649	GIGYFVLOPKVVDGTPCSPDSTSVYCVQGVQCVKAGCDRIIDSKKFDKCGVCGGNGSTCK	708
Qy	684	KVTGLFTKPMHWGNFWPAIPAGAGSIDIRQRYKGLIGDDNYLALKNISQGYLLNGHFVY	743
Db	709	KISGVSYSKAPGYHDIITIPGTAINIEVKQRNQRSGRNSGFLAIKADGTYILNGDYTL	768
Qy	744	SAVERDLVVKGLLRYSGTGTAVESLQASRPILBPLTVEVLVSGKMTTPPRVYSFYLKPE	803
Db	769	STLEQDINMYKVVLYSGSSAALERISFSPKPLTQVLTVGNALRPKIRYTFVFKK	828
Qy	804	PREDKSSHKDPGRGSPVLSHNSVLSLSNQVEQDDRPAPRWAGSWGSPCASGSLQKRA	863
Db	829	-----KES-----FNAIPFS-----ANVIEWGECSKSELGQRL	861
Qy	864	VDCRGAGQRTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCSKSGRGQRSLKC	920
Db	862	VECEDINGQ---PASECAKEVPASTRCPADHPCPQWQLGEWSSCKTKCGYKKRSLAC	918
Qy	921	VGHGRLRLARDQCNLHRKQOE-LDFCVLRPC	950
Db	919	LSHDGGVLSHESCDPLKPKHFIDFCTMAEC	949
RESULT 7			
AAB73549			
ID	AAB73549 standard; Protein; 950 AA.		
XX	AAB73549;		
XX			
XX			
DT	07-AUG-2001 (first entry)		
DE	Human ADAM-type metalloprotease MDT54, SEQ ID NO:4.		
XX			
KW	Human; MDT54; ADAM-type metalloprotease; drug screening;		
KW	A Disintegrin And Metalloprotease; cancer; arthritis.		
OS	Homo sapiens.		
XX			
PN	JP2001017183-A.		
XX			
PD	23-JAN-2001.		
XX			
PF	09-JUL-1999; 99JP-0196584.		
XX			
PR	09-JUL-1999; 99JP-0196584.		
XX			
PA	(YAMA) YAMANOUCHI PHARM CO LTD.		
DR	WPI; 2001-275950/29.		
XX			
DR	N-PSDB; AAB20224.		
PT	A new metal protease and its preparation for use as an anti-cancer and		

PT anti-arthritis therapeutic -

PS Claim 1; Page 12-14; 22pp; Japanese.

CC The invention relates to the novel human ADAM (A Disintegrin And
CC Metalloprotease)-type metalloproteases MDTs4 (AAB3549) and MDTs5
CC (AAB3550). The metalloproteases can be used for the treatment of
CC cancers and arthritis. The invention also relates to the genes encoding
CC MDTs4 and MDTs5, vectors and host cells containing the MDTs4 or MDTs5
CC genes, the recombinant production of MDTs4 and MDTs5, and antibody
CC specific for MDTs4 or MDTs5, and methods of screening for compounds
CC which modulate the activity of MDTs4 and/or MDTs5. The present
CC sequence represents human MDTs4.

XX Sequence 950 AA;

Query Match 48.08; Score 2480.5; DB 22; Length 950;
Best Local Similarity 48.99; Pred. No. 6.8e-184;
Matches 485; Conservative 154; Mismatches 251; Indels 101; Gaps 24;

QY 1 MLLGILTFAPAGTAGGFEPEVYVPIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
DB 19 LLLAAALLAVSDALGRSEDEELVP-ELE-----RAP---GHGTTLRRLHAF 64
QY 61 QEDFYLHTTPAQFLAPAFSTEHLG-----VPLQGLTGSSDLRRCFYSGDVNAEPDSF 114
DB 65 DQQLDLRLPDSFLAPGFTLQNGVRKSGSETPLP-----ETDLAHCFYSTVNGDPSSA 119
QY 115 AAVSLCGGLGAFYRGAEYISPLPNAS---APAAQNSOGA---HLIQ---RRGVPGG 165
DB 120 AALSCEGVRGAYLLGAEYFIQPLPAASERLATAPEKPPAPLOPHLLRRNRQGDVGG 179
QY 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPYKPRRAGFSGSRSSRS 207
DB 180 TCGVVDDPRPTGKAETDEDEGEDEGPQWS-----PODPAQGVGP-TCGTS 230
QY 208 GRARFVSIPRYVETLVVADSNVYKFGADLEHYLLLTATAARLYRHPSTLNPINIVVY 267
DB 231 IRKKRFVSSHRYVETMLVADQSMAEFGHGLKHYLLTFLFVAARLYRHPSTIRNSVLVVY 290
QY 268 KVLILDRDSDGPKVTGNAALTNRFCWAKKLVKSPKHPEYWDTAILFTQDLGATTC 327
DB 291 KILVHDEQKGPVTSNAALTNRFCWQKHNPSPDRDAEHYDTALFTQDLGSGQTC 350
QY 328 DTLMADVMTMCDPKRSVIEDDGLPSAFTTAHELGHVFMNPHDNVYKCEVEYFGKLRAN 387
DB 351 DTLMADVMTVCDPSRSCSVIEDDGLQAAFTTAHELGHVFMNPHDDAKQCASLNGVQDS 410
QY 388 HMMSPTLIQIDRANPWSACSNAITDPLDSGHGCLLDQPSKPLSLPEDLPGLASYTLISQQ 447
DB 411 HMMASMLNLDHSQFSPSCSAYMITSFLDNGHGECMLMDKFNQFIQLPGLDLPSTYDANRQ 470
QY 448 CELAFGVGSKPCP-YMOYCTKLMWCTGKAKGMVQCTRHFPWADGTSCEGKCLKLGACVE 506
DB 471 CQFTFGEDSKHCPDAATCSTLWCTGTSGGVLCVQTHFPWADGTSCEGKWCINGKCVN 530
QY 507 RHNLNKH---RVDGSNAKWPYGPCSTCGGVLQARRQCTNPTPANGKYCYGVRVYKR 563
DB 531 KTD-RKHFDTPFHGSGWGMWPGWDCSRCTCGGVQVYTMRECDNVPKNGKYCEGKRVYR 589
QY 564 SCNLEPCPSSAGSKSFREEOCEAPNGYNHSTNRLTAVAMVPKYSGVSPDKCLICRAN 623
DB 590 SCNLEDCPDN-NGKTFREOCEAINEFSKSFSGGPAVEWIPKYAGVSPDKRCKLICQAK 648
QY 624 GTGYFYVLAPKVDGTLCSPDSTSVYQVQKCIKAGCDNGLGSKRFRDKCGVCGDNKSK 683
DB 649 GIGYFFVLQPKVVDGTPCSPDSTSVYQVQCVKRGCDRIIDSKKPKDKCGVCGNGSTCK 708
QY 684 KVTGLTFKPHGYNFVVAIPAGASSIDIRORYGKGLIGDDNYIALKNQSKYLLNGHFVV 743
DB 709 KISGSVTSAPGYHDIITPTGATNIEVKORNGSRNNGSFLAIKAAADGTIILNGDYTL 768
QY 744 SAVERDLVWKSLLRYSSTGTAVESLQASRPILPTVLESVGKMTPPRVRSFYLPKE 803

DB 769 STLEQDIMYKGVLYRSGSAAALERSPSLKEPTLIQVLTVGNALRPAIKYTYVKKK 828
QY 804 PREDKSSHPKDPGRPSVYLHNSVLSLSNQVEQDPDRPPARVAGSWGPCSCSAGSLQKRA 863
DB 829 -----KES-----FNAITFS-----AAWIEEMGECSKSCGLGWORRL 861
QY 864 VDCRGSAGORTVPACDAAH--RPEVTOACGE-PCPTWELSAWSPSCSKSCGRGFORSKLC 920
DB 862 VECRDINGQ---PASECAKEVKKPASTRPCADHPCPOWLGEWSSCKTCGKGYKRSKLC 918
QY 921 VGHGGRLARDQCNLHRKPOE-LDFCVLRPC 950
DB 919 LSHDGGVLSHESCDPLKPKHFDICTMAEC 949
RESULT 8
AAB50002
ID AAB50002 standard; Protein; 950 AA.
XX
AC AAB50002;
XX
DT 19-MAR-2001 (first entry)
XX
DE Human METH1.
XX
KW Human; METH1; metalloprotease; thrombospondin; angiogenesis inhibition;
KW cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischaemic limb angiogenesis; Osler-Webber syndrome;
KW plaque neovascularisation; telangiectasia; haemophilic joint;
KW angiofibroma; fibromuscular dysplasia; wound granulation;
KW Crohn's disease; atherosclerosis; birth control.
XX
OS Homo sapiens.
XX
PN WO2000071577-A1.
XX
PD 30-NOV-2000.
XX
PF 25-MAY-2000; 2000WO-US14462.
XX
PR 25-MAY-1999; 99US-0318208.
PR 20-JUL-1999; 99US-0144882.
PR 10-AUG-1999; 99US-0147823.
PR 13-AUG-1999; 99US-0373658.
PR 22-DEC-1999; 99US-0171503.
PR 22-FEB-2000; 2000US-0183792.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (SMK) SMITHKLINE BEECHAM CORP.
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUBE/) RUBEN S M.
PA (JONA/) JONAK Z L.
PA (TRUL/) TRULLI S H.
PA (FORN/) FORNWALD J A.
PA (TERR/) TERRETT J A.
XX
PI IrueLa-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
PI Fornwald JA, Terrett JA;
XX
XX WPI; 2001-025136/03.
DR N-PSDB; AAC90057.
XX
XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
PT rheumatoid arthritis and psoriasis -
XX

Claim 15; Fig 1; 768pp; English.

PS The present sequence is human METH1 (ME for metalloprotease and TH for thrombospondin). METH1 can be used for inhibiting angiogenesis in an individual, and for treating cancer, benign tumours, an ocular angiogenic disease, rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis, vasculogenesis, granulations, hypertrophic scars, nonunion fractures, scleroderma, trachoma, vascular adhesions, myocardial angiogenesis, coronary collaterals, cerebral collaterals, arteriovenous malformations, ischaemic limb angiogenesis, Osler-Webber syndrome, plaque neovascularisation, telangiectasia, haemophilic joints, angiofibroma, CC fibromuscular dysplasia, wound granulation, Crohn's disease or CC atherosclerosis. METH1 can also be used in birth control. METH1 can also be used in diagnostic methods for the prognosis of cancer.

SQ. Sequence 950 AA;

Query Match 48.0%; Score 2480.5; DB 22; Length 950;

Best Local Similarity 48.9%; Pred. No. 6.8e-184; Mismatches 251; Indels 101; Gaps 24;

Matches 485; Conservative 154; Mismatches 251; Indels 101; Gaps 24;

QY 1 MLLGLITLAFAGTAGGFEPEVVPVIRLDPDINGRRYWRGPDSDGDOGLFOITAF 60
 DB LLLLAALLAVSDALGRPSDEDELVP-ELE-----RAP---GHGTRRLHAF 64
 QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGSSDLRRCFYSGDVNAEPDSF 114
 DB DQDLDELRLPDSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCYSGTVNGDPSSA 119
 QY 115 AAVSLCGLRGARGYGAEVISPLPNAS---APAQRNSQGA---HLLQ---RRGVPG 165
 DB AALSCEGVRCGAYLLGEAYFIOLPAAERLATAAPGKPPAPLOFHLRRNRQGDVG 179
 QY 166 PSG-----DPTSRC-----GVASG--WNPAITRALDPYKPRRAGEGSRSRRS 207
 DB TCGVVDDEPRPTGAETDEDETEGEDEGPQMS-----FQDPALQGVGP-TGTGS 230
 QY 208 GRAKRFVSIPIRYVETLVVADSMVKFHGADLEHYLLTLATAARLYRHPISILNPINIVV 267
 DB IRKRFVSSHRYVETMLVADQSMARFHSGLKHYLLTLFSAARLYKHPISIRNSILVVV 290
 QY 268 KVLRLDRDSGPKVTGNAALTLNFCNAOKLKNVSKDHPYWDYTLTFQDLCGATTC 327
 DB KILVHDEQKGPVTSNAALTLNFCNWKQHPNDPDRDAEHYDTALTFQDLCGQTC 350
 QY 328 DTLMADVGMCDPKRSCVIEDDGLPSAFTTAHELGHVFNMPHDNPKVCEEVFGKLRAN 387
 DB DTLMADVGMCDPKRSCVIEDDGLQAAFTTAHELGHVFNMPHDNPKVCEEVFGKLRAN 410
 QY 388 HMSPTLIQIDRANPWSACSAIITDFLDGHDGCLLDQPSKPISLPDLPGASYTLSSQ 447
 DB HMASMLSLNDHSQPSWSPSCSAYMITSFLDNGHGECLMDKPNPQLPGLDLPGTSDANRQ 470
 QY 448 CELAFGVGSRPCP-YMYCYTKLACTGAKQMVCOQTHFFWADGTSCGEGKLLKACVE 506
 DB CQFTFGEDSKHCPDAASTCTSLWCTGTSGVGLVCOQTHFFWADGTSCGEGKLLKACVE 530
 QY 507 RHNLNKH---RVDGSAKWDPYPCSTCGGVLARQCTNPTPANGKYCEGVRYKYR 563
 DB KTD-RKHFDTPFHGSGWGWGDCSTCGGVQVYTHRECDNVPNGKYCEGVRYKYR 589
 QY 564 SCNLEPCPSSASGKFRFEOCAFPNGYNHSTNRLTLAVAWPKYSGVSPKDKCLICRAN 623
 DB SCNLEDCPDN-NGKTFREOCEAHNEFSKASFGSGPAAVEIPKAYGSPKDKCLICQAK 648
 QY 624 GTGYFYVLAPVVDGTLCSDDSTSVCGVKCIKAGCDNGLSKRRDKCGCGDNKSK 683
 DB GIGYFFVLQPKVVDGTPCSPDSTSVCGQCVKAGCDRIIDSKKKDKCGCGGNGSTCK 708
 QY 684 KVTGLFKPMHGVNFVAIPAGASSIDIRGVYKGLIGDNDNYLALKNQSGKYLINGHFVV 743
 DB KISGVSATKPGYHDIITPTGATNIEVKQNRQGRSNGSFLAIAKAADGTIILNGDYTL 768

QY 744 SAVERDLVVKGSLRLRYSGTGTAVESLQASRILEPLTVEVLSVCKMTPPRVRYSYFLPKE 803
 DB 769 STLEODIMYKGVRLRYSGSSAALERISPLKEPLTQVLTGVNLRPKIKTYFYFVKKK 828
 QY 804 PREDKSSHPKDPGRGFSVLHNSVLSNQVEQDDPPARWVAGSWGPCSCSGSLQKRA 863
 DB 829 ----KES-----FNAIPTFS-----AWVIEWGECSSKCELGWQRRRL 861
 QY 864 VDCRSGAGQRTVPACDAAH--RPVETQACGF-PCPTWELSAWSPCKSCGGRFQRRSLKC 920
 DB 862 VECRDINGQ---PASECAKEVKPASTRCADHPCQWQWLGWSSCKTCGKGKYSRLKC 918
 QY 921 VGHGGRLLARDQCNLHRKQPE-LDFCVLRPC 950
 DB 919 LSHDGVLSHESCDPLKKPKHFIDFCTMAEC 949

RESULT 9

AAW80285

ID AAW80285 standard; Protein; 967 AA.

XX AAW80285;

DT 19-JAN-1999 (first entry)

XX Human integrin ligand polypeptide ITGL-TSP.

XX ITGL-TSP; integrin ligand; angiogenic disease; cancer; atherosclerosis;
 KW chronic inflammatory disorder; rheumatoid arthritis; tissue remodeling;
 KW macular degeneration; diabetic retinopathy; Alzheimer's disease; human;
 KW restenosis.

OS Homo sapiens.

XX EP874050-A2.

XX 28-OCT-1998.

XX 27-JAN-1998; 98EP-0300575.

XX 24-APR-1997; 97US-0845496.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PA (SMIK) SMITHKLINE BEECHAM PLC.

XX Fronwald JA, Hastings GA, Jonak ZL, Terrett JA;

PI Trulli SH;

XX WPI; 1998-54643/47.

DR N-PSDB; AAV66508.

XX DNA encoding integrin ligand polypeptide ITGL-TSP - used to treat
 PT angiogenic diseases, restenosis, Alzheimer's disease and in tissue
 PT remodeling

XX Claim 11; Pages 6-9; 24pp; English.

XX This represents a human integrin ligand polypeptide ITGL-TSP. ITGL-TSP
 CC polypeptides can be used in the treatment of angiogenic diseases such as
 CC cancer, cancer metastasis, chronic inflammatory disorders, rheumatoid
 CC arthritis, atherosclerosis, macular degeneration or diabetic retinopathy,
 CC restenosis, Alzheimer's disease and tissue remodeling. They can be used
 CC to treat a subject in need of enhanced activity or expression of the
 CC ITGL-TSP polypeptide.

XX Sequence 967 AA;

Query Match 48.0%; Score 2480.5; DB 19; Length 967;

Best Local Similarity 48.9%; Pred. No. 7e-184;

Matches 485; Conservative 154; Mismatches 251; Indels 101; Gaps 24;

QY 1 MLLGLITLAFAGTAGGFEPEVVPVIRLDPDINGRRYWRGPDSDGDOGLFOITAF 60

Db 36 LLLAAALVSDALGRSEDEELVWP-ELE-----RAP-----GHGTRRLRHAF 81
QY 61 QEDFYLHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSDLRRCFYSGDVNAEPDSF 114
Db 82 DQDLLELRPDSSFLAPGFTLQNGVRKSGSETPLP-----ETDLAHCFYSGTVNGDPSSA 136
QY 115 AAVSLCGGLRGAFYRGAEYVTSPLPNAS-----APAAQRNSOGA-----HLQO--BRGYPGG 165
Db 137 AALSCEGVGAFFYLLGAYFIQPLPAASERLATAAPGEKPPALQFHLRLNRQGVGG 196
QY 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPYKPRRAGFSGSRRRS 207
Db 197 TCGVYDDPRPTGKAETDEDEGEDEGPQWS-----PQDPALQGVGP-TGTGS 247
QY 208 GRAKEFVSIPRYVETLVVADSMVKFHGADLEHVLTLTLLATAARLYRHPSTILNINIVV 267
Db 248 IKKRFVSSHRYVETLVVADSMVAFHGGSLKHLVLLTFSVAARLYKHPSTIRNSVSLVV 307
QY 268 KYLLLRDRDGGPKVTGNAALTLRNFCAMQKKLVSKVSDKPEYWDTAILFTRODLCGATTC 327
Db 308 KILVTHDEQKGPVTSNAALTLRNFCNWKQKHPSPDRDAEHYDTAILFTRODLCGATTC 367
QY 328 DTGLNADVGTWCDPKRSCSVTEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEEFGKLRAN 387
Db 368 DTGLNADVGTWCDPKRSCSVTEDDGLQAAFTTAHELGHVFNMPHDNVKVCVEEFGKLRAN 427
QY 388 HMMSPFTLQIDRANPWSACSAITITDLDGSHGDCLLDPSKPSLSPEDLPASVYTLISQO 447
Db 428 HMMASLNLDSHPWSPSCSAITITDLDGSHGDCLLDPSKPSLSPEDLPASVYTLISQO 487
QY 448 CELAFGVSGKPCP-YMQVCTKLWCTGKAKQMVCTRHFPWADGTSCGEGKCLKLGACVE 506
Db 488 CQFTFGEDSKPCPAAASTCSTLWCTGSGVVLVQTKHFPWADGTSCGEGKCLKLGACVE 547
QY 507 RHNLNKH---RVDGSWAKWDPYPCSTCGGVQOLARQCTNPTTPANGKYCEGVKRYR 563
Db 548 KTD-RKHEDTFFHGSWGMWGPWCGDCSTCGGVQYTTRECDNPVPKNGKYCEGVKRYR 606
QY 564 SCNLEPCPSSASGRFEEQCEAFNGYNHSTNRLTLAVAWVPKYSGVSPRDKCLICRAN 623
Db 607 SCNLEDCPDN-NGKTFREEQCEAHNEFSKASFGSGPAVEWIPKYAGVSPDRCKLICQAK 665
QY 624 GTGYFYVLAPKVDGTLCPDSTSVCGQKIKAGCDGNLGSKKRFDKCGVCGGDNKSK 683
Db 666 GIGYFFVLQPKVDGTFCPSDSTSVCGQKIKAGCDRIIDSKKFDKCGVCGGNGSTCK 725
QY 684 KVTGLTKPMHGYNFVVAIPAGASSIDIRORYKGLIGDNDNYLALKNSQKYLINGHFVV 743
Db 726 KISGVTSAPGYHDITIPGATNIEVKQNRQSGNRNNGSFGLAKAADGTYLINGDYTL 785
QY 744 SAVERDLVVGSLRLYSGTGTAVESLQASRPILFLETVESVLSVKMTPPRVRYSFYLPKE 803
Db 786 STLEQDINMYKGVLRYSGSSAALRIRSFSPLEPLTIQVLTGNALRPKIKTYFVKKK 845
QY 804 PREDKSSHPKDPGCVLHNSVLSLQVQPDPRPARVAGSWGPCSCSGSLQKRA 863
Db 846 -----KES-----FNAIPTFS-----AWVIEWEGSCSKSCELGWQRL 878
QY 864 VDCRGSAGQRTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCKSGCGRFORSLKC 920
Db 879 VECRDINGQ---PASECAKEVKPASTRPCADHPCPQWLGENWSCSTCKGKYKRSLLKC 935
QY 921 VGHGGRLLARDQCNLHKKPQB-LDFCVLRPC 950
Db 936 LSHDGGVLSHESCDPLKPKHFIDFCTMAEC 966

RESULT 10
AAB50011
ID AAB50011 standard; Protein: 968 AA.
XX
AC AAB50011;

XX
DT 19-MAR-2001 (first entry)
XX DE Protein; SEQ ID 125.
XX DE
KW Human; METH1: metalloprotease; thrombospondin; angiogenesis inhibition;
KW cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischaemic limb angiogenesis; Osler-Webber syndrome;
KW plaque neovascularisation; telangiectasia; haemophilic joint;
KW angiofibroma; fibromuscular dysplasia; wound granulation;
KW Crohn's disease; atherosclerosis; birth control.
XX Homo sapiens.
XX WO200071577-A1.
XX 30-NOV-2000.
XX 25-MAY-2000; 2000WO-US14462.
XX 25-MAY-1999; 99US-0318208.
XX 20-JUL-1999; 99US-0144882.
XX 10-AUG-1999; 99US-0147823.
XX 13-AUG-1999; 99US-0373658.
XX 22-DEC-1999; 99US-0171503.
XX 22-FEB-2000; 2000US-0183792.
XX (HUMA-) HUMAN GENOME SCI INC.
XX (SMK-) SMITHKLINE BEECHAM CORP.
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX (IRUE/) IRUELA-ARISPE L.
XX (HAST/) HASTINGS G A.
XX (RUBE/) RUBEN S M.
XX (JONA/) JONAK Z L.
XX (TRUL/) TRULLI S H.
XX (FORN/) FORNWALD J A.
XX (TERR/) TERRETT J A.
XX Iruela-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
PI Fornwald JA, Terrett JA;
XX WPI; 2001-025136/03.
XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
PT rheumatoid arthritis and psoriasis -
XX Claim 15; Pages 759-763; 768pp; English.
XX The present invention relates to human METH1 and METH2 (ME for
CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
CC METH can be used for inhibiting angiogenesis in an individual, and for
CC treating cancer, benign tumours, an ocular angiogenic disease,
CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
CC fibromuscular dysplasia, wound granulation, Crohn's disease or
CC atherosclerosis. METH can also be used in birth control. METH can also
CC be used in diagnostic methods for the prognosis of cancer. The present
CC sequence is a protein isolated in the present invention.
XX Sequence 968 AA;

Query Match 48.0%; Score 2480.5; DB 22; Length 968;
Best Local Similarity 48.9%; Pred. No. 7e-184;
Matches 485; Conservative 154; Mismatches 251; Indels 101; Gaps 24;

QY 1 MLLIGILTLAFAAGTAGGFEPEPEVIRLDPDINGRRYRWGRPEDSGDGLIFQITAF 60
 DB 37 LLLLAALLAVSALGRPSDEELVVP-ELE-----RAP---GHGTRRLHAF 82
 QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRCFYSGDVNAPDSF 114
 DB 83 DQQLDELRLPDSFLAPGFTLQNGRKSSETPLP-----ETDLAHCFTYSGTVNGDPSSA 137
 QY 115 AAVSLCGGLRGAFYGAAYVISPFPNAS---APAAQRNSOGA-----HLLQ---RRGVPGG 165
 DB 138 AALSCEGVRYGAYFIQPLPAASERLATAAPGKPPAPLOFHLRLRRNRQGDVG 197
 QY 166 PSG-----DPTSRC-----GVASG--WNPAIRLALDPYKPRRAGFGESRRRS 207
 DB 198 TCGVVDDEPRPTGKAETDEDETEDEGEPQWS-----PQDPALQGVGP-TGTGS 248
 QY 208 GRAKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPISILNPIV 267
 DB 249 IRKRFVSSHRYVETMLVADQSMAEFHGSLGHYLLTLFSAARLYKHPISIRNSVLV 308
 QY 268 KVLRLDRDGGPKVTGNAALTNRNFCAMOKKLNKVSXKHPEYWDTAILFTRODLCGATTC 327
 DB 309 KILVIHQKPEVTSNAALTNRNFCAMOKKLNKVSXKHPEYWDTAILFTRODLCGATTC 368
 QY 328 DTLCGADVGTMCOPKRCSCVIEDDGLPFAFTTAHELGHVENMPHDNVKVEEYFGKLRAN 387
 DB 369 DTLCGADVGTCDPSRSCVIEDDGLQAAFTTAHELGHVENMPHDNVKVEEYFGKLRAN 428
 QY 388 HMSPTLIQIDRANPWSACSAITPDLSDHGDCLLDQPSKIPISIPEDLPASYSYLSQ 447
 DB 429 HMASMLNLDHSPQPCSAAYMITSFLDNGHCECLMKPQNPQLPGLPDTSYDANRQ 488
 QY 448 CELAFVGSKPCP-YMOCYTKLACTKAKQMVCONPFPWADGTCGEGKCLKKACVE 506
 DB 489 CQFTGEDSKHCPDPAATCTSLMCTGTSGVGVLCQTHKFPWADGTCGEGKCLKKACVE 548
 QY 507 RHNLNKH---RVDSWAKWDPYPCSTCGGVQGLARRCOTNTPANGGKYCEGVRYK 563
 DB 549 KTD-RKHFDTPFHGSGWGMWPGWDCSRTCGGVQVYTMRECDNPVKNNGKYCEGVRYK 607
 QY 564 SCNLEPCSSAGSKFPEOCEAFNGYHNSTNRLTLAVAVPKYSPVSDKCLICRAN 623
 DB 608 SCNLEDCPDN-NGKTFREQCEAHNEFSKASFGSGPAVEPIKAYGVSPKDRCKLICQAK 666
 QY 624 GTGYFYVLAQVVDGTLCSDPSTSVQVQKICIRAGCDNLGSKRRFDKCGVCGDNKSK 683
 DB 667 GIGYFVFLQPKVVDGTCSPDSTSVQVQKICIRAGCDRIIDSKKFKDKCGVCGDNKSK 726
 QY 684 KVTGLTKPMHGYNVVAIPAGASSIDIRORYKGLIGDDNYLAKNSQGYLLNGHFVV 743
 DB 727 KISGVTSAKPGYHDIITPTGATNIEVKORNGSRNNGSFLAIAAADGTYILNGDYL 786
 QY 744 SAVERDLVVGSLRYSGTAVESLOASRPILEPLVEVLSVKMTPPVRYSFYLPKE 803
 DB 787 STLEQDLMKGVVLYSGSSAALRIKSFLEPLQIQLVTGNALRKIKITFYVKKK 846
 QY 804 PREDKSHPKDPGRPSVLHNSVLSNQVEQDPRPARRVAGSWGSPCASGSGSLQKRA 863
 DB 847 ---KES-----FNAIPFS-----ANVIEWGCSSKSELGWRRL 879
 QY 864 VDCRGSAQRTVPACDAAH---RPVETQACGE-PCPTWELSAWSPCSKSGRGQRRLKC 920
 DB 880 VECRDINGQ---PASECAKEKVPACDHPCPQWOLGEWSSCKTKCGKYKRSKLC 936
 QY 921 VGHGGLLARDQCNLHRRKPOE-LDFCVLRLPC 950
 DB 937 LSHDGGVLSHESCDPLKPKHFIDFCTMAEC 967

RESULT 11
 AAY04142
 ID AAY04142 standard; Protein; 967 AA.

XX AAY04142;
 AC 15-JUN-1999 (first entry)
 DT Human Tango-71 protein.
 DE Human; Tango-71; Tango-73; Tango-74; Tango-76; Tango-83; diagnosis;
 KW detection.
 KW Homo sapiens.
 OS WO9907850-A1.
 PN 18-FEB-1999.
 PD 06-AUG-1998; 98WO-US16502.
 PF 05-SEP-1997; 97US-0058108.
 PR 06-AUG-1997; 97US-0054966.
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 XX Goodearl ADJ, Holtzman DA;
 PI WPI; 1999-167426/14.
 DR N-PSDB; AAX19955.
 XX New TANGO polypeptides and nucleic acids encoding them - useful as
 PT diagnostic agents and for treating disorders caused by aberrant
 PT expression of TANGO
 XX Claim 8; Fig 1; 84pp; English.
 PS The present sequence represents human Tango-71. Tango polypeptides are
 CC useful for identifying compounds which bind the polypeptide via direct
 CC binding, competition binding assays or Tango-71, -73, -74, 76 or -83-
 CC mediated signal transduction. Tango polypeptides are also useful for
 CC identifying modulating compounds by determining effect on Tango activity.
 CC Tango polypeptides and nucleic acids are useful for diagnosing diseases
 CC related to aberrant expression of Tango, and Tango polypeptides are
 CC useful for raising antibodies which can be used in diagnostic assays for
 CC detection of Tango, and also for generating anti-idiotypic antibodies for
 CC prevention and protection.
 XX SQ Sequence 967 AA;
 Query Match 48.0%; Score 2479.5; DB 20; Length 967;
 Best Local Similarity 48.9%; Pred. No. 8.4e-184;
 Matches 495; Conservative 154; Mismatches 251; Indels 101; Gaps 24;
 QY 1 MLLIGILTLAFAAGTAGGFEPEPEVIRLDPDINGRRYRWGRPEDSGDGLIFQITAF 60
 DB 36 LLLLAALLAVSALGRPSDEELVVP-ELE-----RAP---GHGTRRLHAF 81
 QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRCFYSGDVNAPDSF 114
 DB 82 DQQLDELRLPDSFLAPGFTLQNGRKSSETPLP-----ETDLAHCFTYSGTVNGDPSSA 136
 QY 115 AAVSLCGGLRGAFYGAAYVISPFPNAS---APAAQRNSOGA-----HLLQ---RRGVPGG 165
 DB 137 AALSCEGVRYGAYFIQPLPAASERLATAAPGKPPAPLOFHLRLRRNRQGDVG 196
 QY 166 PSG-----DPTSRC-----GVASG--WNPAIRLALDPYKPRRAGFGESRRRS 207
 DB 197 TCGVVDDEPRPTGKAETDEDETEDEGEPQWS-----PQDPALQGVGP-TGTGS 247
 QY 208 GRAKRFVSIPIRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPISILNPIV 267
 DB 248 IRKRFVSSHRYVETMLVADQSMAEFHGSLGHYLLTLFSAARLYKHPISIRNSVLV 307
 QY 268 KVLRLDRDGGPKVTGNAALTNRNFCAMOKKLNKVSXKHPEYWDTAILFTRODLCGATTC 327

Db 308 KILVIHQEGPEVTSNAALTLRNFNCNQKHQHPNPPSDRDAEHYDTAILTRQDLGSGQTC 367

QY 328 DTLMADVGTMCDPKRSRCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRAN 387

Db 368 DTLMADVGTCDPSCRSVIEDDGLQAFHTAHELGHVFNPHDDAKQOCASLNGVNDSD 427

QY 388 HMMSPTLTIOIDRANPWSACSAIIITFDLSDGHGDCLLDPKSPISLSPEDLPGASVYTLSSQ 447

Db 428 HMMASMLSLNDHSQSPSCSAVMYITSLDNGHGECLMDKPQNPIQLPGDLPGTSYDANRQ 487

QY 448 CELAFGVGSKPCP-YMQVCTKLWCTGKAKGQWCVOTRFPFWADTSCGEGKLCIKGACVE 506

Db 488 COFTGEDSKHCPDAASTCTLWCTGTSGGVLCVOTKHPWADTSCGEGKWCINGKCVN 547

QY 507 RHNLANH---RVDSWAKWDPTGCSRTCCGGVQLARQCNTPTPANGKYCEGVRVRYR 563

Db 548 KTD-RKHEDTTPHSGWGMWPGWDCSRTCCGGVQYTWECNDPVPKNGKYCEGKRVRYR 606

QY 564 SCNLEPCPSSASGKSFREOCEAFNGYHNHSTNRLTLAVAWPKYSGVSPDKCKLICRAN 623

Db 607 SCNLEDCPDN-NGKTFREOCEAHNEFSKASFGSPAVEMIPKYAGVSPDKCKLICQAK 665

QY 624 GTGYEVLAPKVDGTLCSPDSTSCVCGKCIKAGCDNLGSKKREDKCGVCGGDNKSKC 683

Db 666 GIGYEFVLQKVDGTPCSPDSTSCVCGQCVKAGCDRIIDSKKFKDKCGVCGGNGSTCK 725

QY 684 KVTGLTFPMHGYNFWVAIPAGASSIDIRQRYGKGLIGDDNLYALKNSQKRYLLNGHFV 743

Db 726 KISGVTSAKPYHDIITPTGATNIEVKQNRQSRNNGSFLAIIKAADGTIYILNGDYTL 785

QY 744 SAVERDLVVKSLRYCTGTAVESLQASRIEPLAVEVLVSQKMTPPRVKYSFYLPKE 803

Db 786 STLEDIMYKGVLYRSGSSAALRISFSLKEPLTQVLTGNALRPKIKYTFYVKK 845

QY 804 PREDKSHPKDPGRPSVLHNSVLNSQVEQDPRPARWAGVSGPCSCGSGLOKRA 863

Db 846 ---KES-----FNAIPFS-----AWIEWGCSKSCELGWQRRL 878

QY 864 VDCRGSAGQRTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCSKCGRFGORSLKC 920

Db 879 VECRDINGQ---PASECAKEVKPASTRPCADHPQCPQMLGWSWSSCTCGKYKRSKLC 935

QY 921 VGHGRLILARDQCNLHRKPQ-BLDFCVLRPC 950

Db 936 LSHDGGVLSHESCDPLKPKHFDICTMAEC 966

RESULT 12

AAW78189

XX AAW78189 standard; Protein; 967 AA.

XX AC AAW78189;

XX DT 13-APR-1999 (first entry)

XX DE Human secreted protein encoded by gene 64 clone HOUQ17.

XX KW Human; secreted protein; fusion protein; gene therapy; protein therapy; diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia; developmental abnormality; foetal deficiency; blood; allergy; renal; immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma; inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS; cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus; osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion; endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Misc-difference 40

FT /label= unknown

FT Misc-difference 45

FT /label= unknown

FT Misc-difference 169 /label= unknown

FT Misc-difference 293 /label= unknown

FT Misc-difference 297 /label= unknown

FT Misc-difference 557 /label= unknown

XX

PN W09856804-AL.

XX

PD 17-DEC-1998.

XX

PF 11-JUN-1998; 98WO-US12125.

XX

PR 02-OCT-1997; 97US-0061060.

PR 13-JUN-1997; 97US-0049547.

PR 13-JUN-1997; 97US-0049548.

PR 13-JUN-1997; 97US-0049549.

PR 13-JUN-1997; 97US-0049550.

PR 13-JUN-1997; 97US-0049606.

PR 13-JUN-1997; 97US-0049607.

PR 13-JUN-1997; 97US-0049608.

PR 13-JUN-1997; 97US-0049609.

PR 13-JUN-1997; 97US-0049610.

PR 13-JUN-1997; 97US-0049611.

PR 13-JUN-1997; 97US-0050566.

PR 13-JUN-1997; 97US-0050901.

PR 13-JUN-1997; 97US-0052989.

PR 08-JUL-1997; 97US-0051919.

PR 18-AUG-1997; 97US-0055984.

PR 12-SEP-1997; 97US-0058665.

PR 12-SEP-1997; 97US-0058668.

PR 12-SEP-1997; 97US-0058669.

PR 12-SEP-1997; 97US-0058750.

PR 12-SEP-1997; 97US-0058971.

PR 12-SEP-1997; 97US-0058972.

PR 12-SEP-1997; 97US-0058975.

PR 02-OCT-1997; 97US-0060834.

PR 02-OCT-1997; 97US-0060841.

PR 02-OCT-1997; 97US-0060844.

PR 02-OCT-1997; 97US-0060865.

PR 02-OCT-1997; 97US-0061059.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Brewer LA, Ebner R, Ferrie AM, Feng P, Greene JM, Lafleur DW;

PI Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Young P;

PI Yu GL;

XX

DR WPI: 1999-080881/07.

DR N-PSDB; AAX04374.

XX

PT New isolated human genes and the secreted polypeptides they encode -

PT useful for diagnosis and treatment of e.g. cancers, neurological

PT disorders, immune diseases, inflammation or blood disorders

XX

PS Claim 11; Page 297-300; 380pp; English.

XX

CC This sequence represents a secreted human protein encoded by the gene

CC clone detailed in the descriptor line. The gene can be used to generate

CC fusion proteins by linking to the gene to a human immunoglobulin Fc

CC portion (e.g. AAX04302) for increasing the stability of the fused

CC protein as compared to the human protein only.

CC The invention relates to 86 novel genes and their fragments (nucleic

CC acid sequences: AAX04311-X04410; amino acid sequences AAW78126-W78225)

CC which are useful for preventing, treating or ameliorating medical

CC conditions e.g. by protein or gene therapy. Also, pathological

CC conditions can be diagnosed by determining the amount of the new

CC polypeptides in a sample or by determining the presence of mutations in

CC the new polynucleotides. Specific uses are described for each of the 86

CC polynucleotides, based on which tissues they are most highly expressed in

CC (see AAX04311 for described uses).

XX	Sequence	967 AA;	
SO	Query Match	47.6%; Score 2458.5; DB 20; Length 967;	
	Best Local Similarity	48.5%; Pred. No. 3.6e-182;	
	Matches	481; Conservative 154; Mismatches 255; Indels 101; Gaps 24;	
QY	1	MLLGILTLAFAGTAGGFFPEREVVPIRLDPDINGRRYYWRGPDSDGQGLIFQITAF	60
DB	36	LLLLAALLXVSDALGRPSDEDELVVP-ELE-----RAP---GHGTRLRKHAF	81
QY	61	QEDFYLHTLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSLRLRCFYSGDVNAEPDSF	114
DB	82	DQDLDELRLPDSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCFYSTGVNGDPSSA	136
QY	115	AAVSLCGGLRGAFGYRGAEYVISPLPNAS---APAAQRNSOGA-----HLIQ--RRGVPGG	165
DB	137	AALSICEGVGAFYLLGEAVFIQPLPAASERLXTAAPGEKPPAPLPQFHLRLNRNRQGDVG	196
QY	166	PSG-----DPTSRC-----GVASG--WNPAILRALDPYKPRRAGFSGESRRSRRS	207
DB	197	TCGVVDDEPRPTGRAETDEDEGTEGDEGPQWS-----FQDPALQGVGP-TGTGS	247
QY	208	GRAKRFVSIPIRYVETLVVADESMVKFAGADLEHYLLTLATAARLYRHPHSILNPINIVVY	267
DB	248	IRKRFVSSHRYVETMLVADQSMAEFHGSLKHYLLTLFVVAARLXKHPXIRNSVSLVVY	307
QY	268	KVLLLRDRDSGPKVTGNAALTLRNFCAWOKLNKYSKHPYWDTAITLFTRODLGGATTC	327
DB	308	KILVIHDEQKGPVTSNAALTLRNFCAWOKLNKYSKHPYWDTAITLFTRODLGGATTC	367
QY	328	DTLGMADVGTWCDPKRSCVTEDDGLPSAFTAHLEHGVFNMPHDNVKCEEVFEGKLRAN	387
DB	368	DTLGMADVGTWCDPKRSCVTEDDGLPSAFTAHLEHGVFNMPHDNVKCEEVFEGKLRAN	427
QY	388	HMSFTLIQIDRANPWSACSAIITFDLSHGDCLLDQPSKPISLPEDLPASLYTSLQQ	447
DB	428	HMASMLSNLDHSPWSPSCSAIMYITSLDNGHGECMLDKPQNPIQLPDLPGTSDYANRQ	487
QY	448	CELAFGVSKPCP-YMQYCTKLWCTGAKGQWVQTRHFWADGTSGCGKCLCLKGACVE	506
DB	498	COFTGEDSKPCDPAASCTSLACTGTSGVLVQCTRHFWADGTSGCGKWCINGKVCXV	547
QY	507	RHNLNKH---RVDSNAKWDPYGPCSRTCGGVQVLARQCTNPTPANGKYCEGVRYKYR	563
DB	548	KTD-RKHEDTFFHSGWMWGPWDCSRTCGGVQVYTWRECDNPPVKNGKYCEGVRYKYR	606
QY	564	SNLEPCSSASGKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCLICRAN	623
DB	607	SNLEDCPDN-NGKTFREOCEAHNEFSKASFGSGPAVENIPKTAGVSPDKRCLICQAK	665
QY	624	GTGYFVVLAPKVDGTLGSPDSTSVCGQKCIKAGCDGNLGSKKRFDKCGVCGGDNKSK	683
DB	666	GIGYFVVLAPKVDGTLGSPDSTSVCGQKCIKAGCDGNLGSKKRFDKCGVCGGDNKSK	725
QY	684	KVTGLFTPMHGYNVVWPAIGASSIDTRQYKGLIGDNNYALKNSQKYLINGHVVV	743
DB	726	KISGVSATKPGYHDIITPTGATNIEVKQNRQNRNNGSFLAIKAADGTIYILNGDYTL	785
QY	744	SAVERDLVVKGLLRYSGTGAVESLQASRPTLEPLTVEVLSVGKMTPPRVRYSYFLPK	803
DB	786	STLEQDQWYKGVVLYSGSSAALERISFSPLEPLTQVLTQVGNALPKIYFYFKKK	845
QY	804	PREDKSSHPKDPGRPSVLHNSVLSLNSQVQPDPRPARWAGVGPSCASCGSLQKRA	863
DB	846	-----RES-----FNAIPTS-----AWVTEWEGCSKSCELGWQRRL	878
QY	864	VDCRSAGORTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCKSCGRGFORRSUKC	920
DB	879	VECRDINGQ---PASECAKEVKASTRPCADHPCPQWLGEWSSCKTCGKGYKRRSLKC	935
QY	921	VGGGRLLARDQCNLHRKPQE-LDFCVLRPC	950

DB	936	LSHDGGVLSHESCDPLKKPKHFIDFTMAEC	966
RESULT	13		
AAV53899			
ID	AAV53899	standard; Protein: 950 AA.	
XX	AC	AAV53899;	
XX	DT	13-MAR-2000 (first entry)	
XX	DE	Amino acid sequence of a murine ADAMTS-1 protein.	
XX	KW	GON-1; metalloproteinase; cell migration; modulator; ADAMTS-1;	
KW	KW	metalloproteinase domain; thrombospondin domain; abnormal cell migration;	
KW	KW	organ shaping; sterility; cancer metastasis.	
XX	OS	Mus sp.	
XX	PN	WO9961656-A2.	
XX	PD	02-DEC-1999.	
XX	PF	28-MAY-1999; 99WO-US11918.	
XX	PR	29-MAY-1998; 98US-0087170.	
PR	PR	13-APR-1999; 99US-0129023.	
XX	XX	(WISC) WISCONSIN ALUMNI RES FOUND.	
PA	XX	Kimble JE, Bielloch RH;	
PI	XX	WPI; 2000-072633/06.	
XX	DR	Identifying modulators of proteins containing metalloprotease and	
XX	PT	thrombospondin domains, potentially useful for controlling cell	
PT	PT	migration and organ shaping	
XX	XX	Disclosure; Fig 1C; 60pp; English.	
PS	XX	The present sequence represents a murine ADAMTS-1 protein. ADAMTS-1 is	
CC	CC	a metalloproteinase. The specification describes another related	
CC	CC	metalloproteinase, a C. elegans GON-1 protein, that lacks a transmembrane	
CC	CC	domain and possesses a predicted metalloproteinase domain between residues	
CC	CC	269-456. In C. elegans hermaphrodites, GON-1 is required for migration of	
CC	CC	two distal tip cells to produce elongated tubes, whereas in males, GON-1	
CC	CC	is required for migration of a single linker cell to produce a single	
CC	CC	elongated tube. The protein is used in the method of the invention. The	
CC	CC	specification describes a method for identifying a modulator of a	
CC	CC	protein that contains a metalloproteinase domain and a thrombospondin	
CC	CC	domain. The method comprises treating a target organism, having a	
CC	CC	developing gonadal cell that is responsive to the protein, with a test	
CC	CC	compound, and determining any change in migration or shape of the cell	
CC	CC	attributable to the test compound. The compounds identified are	
CC	CC	potential therapeutic modulators of abnormal cell migration and organ	
CC	CC	shaping, e.g. for rendering animals (specifically nematodes) sterile	
CC	CC	and for inhibiting cancer metastases.	
XX	XX	Sequence 950 AA;	
QY	1	MLLGILTLAFAGTAGG--FEPERVVVPIRLDPDINGRRYYWRGP-EDSGDGLIFOI	57
DB	20	LLLLASTMLLCARGAHGRETEDELVLP-SLE-----RAPGHDSITTRL--RL	66
QY	58	TAFEDFYLHTLTPDAQFLAPAFSTEHLGVPLOGLTGS-----SDLRRCFYSGDVNA	109
DB	67	DAFQQQLHLKLPDPSGLAPGFTLQTV-----GRSPGSEAHLDPTGDLAHCFYSTGVNG	121
QY	110	EPDSFAVAVSLCGGLRGAFGYRGAEYVISPLPNAS-----APAAQRNSOGA----	161

```
Db 122 DFGSAAALSLCEGVAGFYLOGEFFIOPAGVATERLAPAVPEESSARPQFHILRR- 180
QY 162 VFGGSGDPTSCGVASGWNPAILRALDPYKPRAGFESRR- 206
Db 181 ----RRSGGAKGVND-----DETLPDSRPESQNTNRNOMPVRDPTPOACKP 226
QY 207 ----SGRAKRFVSPRYVETLVVADESVMVKFHGADLEHYLLTLATAARLYRHPSTLNPI 262
Db 227 SGPGRSTRKRFVSSPRYVETMLVADQSMADFHGSLKHYLLTLFVAARFYKHPSTRNSI 286
QY 263 NIIVVKVLLLRDSCPKVTGNAALTNRFCWOKKLNKYSKDHPEYWDTAILFTRODLC 322
Db 287 SLVVVKILVIEQKGPVETSNAAALTNRFCWOKKLNKYSKDHPEYWDTAILFTRODLC 346
QY 323 GATTCDTLGMADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVEFG 382
Db 347 GSHTCDTLGMADVGTVCDPKRSVIEDDG-QAAFTTAHELGHVFNPHDPAKHCASLNG 406
QY 383 KLRANHMSPTLIOIDRANPWSACSAALITDFLDSHGDCLLDOPSKPISLPELPGASY 442
Db 406 VSGDHLKASMLSSLDHSPQSPSAWVTSFLDNGHGECIMDKPQNPRLPSDLPGLTY 465
QY 443 TLSQCELAFGVSGKPCP-YMQYCTKLWCTGKAKGQVCTRFPWADGTSCEGKCLLK 501
Db 466 DANROCCFTFGESKHCPCDAASTCTTLWCTGSGLLVCCCTKHPWADGTSCEGKWCVS 525
QY 502 GACVERHNLKH---RVDGSAKNDPYGCSRTCCGGVQLARRQCTNPTPANGKYCEGV 558
Db 526 GACVKNKTDM-KHFAFVHGSGPMPGDCSRTCCGGVQVQTMRECDNPPVKMGKYCEGK 584
QY 559 RVKYRSCNLEPCSSASGSKFREQCAFNGYHSTNRLTLAVAWPKYSGVSPRDKCL 618
Db 585 RYRYSRNLIEDCPDN-NGKTFREOCEAHNEFSKASFGNEPTVETWPKYAGVSPRDKCL 643
QY 619 ICRANGTYFYVLAKVVDGTLCSFSDTSVCVQKCKIACGDCNGLSKKRFKCGVCGGD 678
Db 644 TCEAKGIGYFVLQPKVVDGTPCSPDSTSVQVQKCKIACGDCNGLSKKRFKCGVCGGD 703
QY 679 NKSCKVTKGLFTKPMHGNVVAIPAGASSIDIRQRYKGLIGDNTYALKNSQGYLLN 738
Db 704 GSTCKKMSGIVTSTPRGYHDIVTIPAGATNIEVKHNRQSRNNGSFLAIRADGTYILN 763
QY 739 GHFVVSVERDLVKGSLRLSGTGAVESLQASRPILPLETVELVSVKMTPPRVYSF 798
Db 764 GNFTSLTEQDITYKGYTLRYSGSAALEIRSFSLPELTTQVLMVGHALRPKIKFY 823
QY 799 YLPKEPREDKSHPKDPRGSPVLSLSNQVEQDDPRPARWAGSWGPCSCSG 858
Db 824 FMKKKTES-----FNAIPTES-----EWIEEWGECSTCGSG 856
QY 859 LQKRAVDCRGSAGQRTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCSKSGRGFOR 915
Db 857 WQRRVVQCRDINGH---PASCAKEVKAPESTRPCADLPCHPHQVGDWSPCSKTCGKGYK 913
QY 916 R-SLKCVGHGRLARDCNLRKPKQE-LDFCVLRPC 950
Db 914 RCTLKCVDHGGVLSNCSDDLKPKKHYIDFCTLTQC 950
```

RESULT 14

```
AAB21265
ID AAB21265 standard; Protein; 896 AA.
XX
XX AAB21265;
XX
XX
DT 23-FEB-2001 (first entry)
XX
XX Mouse metalloproteinase ADAMTS-1.
XX
XX Mouse; ADAMTS-1; metalloproteinase; ADAM;
KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; nontropic; neuroprotective; antiparkinsonian;
```

```
KW cerebroprotective; cytostatic; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.
XX
XX Mus musculus.
XX WO200053774-A2.
XX 14-SEP-2000.
XX 08-MAR-2000; 2000WO-US06237.
XX 08-MAR-1999; 99US-0264585.
XX (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX Kelner GS, Clark M, Maki RA;
XX WPI; 2000-594326/56.
XX
XX Polynucleotide encoding novel members of a disintegrin,
XX metalloproteinase and thrombospondin domain protein family used to
XX prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX Disclosure; Fig 17; 129pp; English.
XX
XX The present sequence is mouse metalloproteinase ADAMTS-1. The
XX ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
XX and Metalloproteinase domain) family. Members of the ADAMTS family
XX contain a thrombospondin domain in addition to the disintegrin and
XX metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
XX useful for the manufacture of medicaments for treating conditions
XX associated with neuroinflammation and/or neurodegeneration, such as
XX Alzheimer's disease, Parkinson's disease and stroke. They are also
XX useful for treating conditions associated with cell proliferation, cell
XX migration, inflammation and/or angiogenesis, such as cancer, arthritis
XX and autoimmune diseases. They can be used to treat patients afflicted
XX with an invasive tumour, a brain tumour or brain injury.
XX
XX Sequence 896 AA:
```

```
Query Match 45.0%; Score 2322; DB 21; Length 896;
Best Local Similarity 48.1%; Pred. No. 1.3e-171;
Matches 453; Conservative 143; Mismatches 236; Indels 110; Gaps 21;
```

```
QY 1 MLLLGILTLAFAGTAGG--PEPEREVVPIRLDPDINGRYYWRGP-EDSGDQGLIFI 57
Db 20 LLLASITMLLCARGAHGRPTDEELVLP-SLE-----RAPGHDSTTTL--RL 66
QY 58 TAFQEDFYHLPTDAQFLAPAFSTEHLGVPLQGLTGS-----SDLRRCFYSGDVNA 109
Db 67 DAFGQQLHLKLOPDSGFLAPGTLQTV-----GRSPGEAOHLDPDGLAHCFYSGTVNG 121
QY 110 EPDSFAAVSLCGLRGAFGYRGAEYVISPLPNAS-----APAAQRNSOGA----HLLQRRG 161
Db 122 DPGSAAALSLCEGVAGFYLOGEFFIOPAGVATERLAPAVPEESSARPQFHILRR- 180
QY 162 VPGGSGDPTSCGVASGWNPAILRALDPYKPRAGFESRR- 206
Db 181 ----RRSGGAKGVND-----DETLPDSRPESQNTNRNOMPVRDPTPOACKP 226
QY 207 ----SGRAKRFVSPRYVETLVVADESVMVKFHGADLEHYLLTLATAARLYRHPSTLNPI 262
Db 227 SGPGRSTRKRFVSSPRYVETMLVADQSMADFHGSLKHYLLTLFVAARFYKHPSTRNSI 286
QY 263 NIIVVKVLLLRDSCPKVTGNAALTNRFCWOKKLNKYSKDHPEYWDTAILFTRODLC 322
Db 287 SLVVVKILVIEQKGPVETSNAAALTNRFCWOKKLNKYSKDHPEYWDTAILFTRODLC 346
QY 323 GATTCDTLGMADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVEFG 382
Db 347 GSHTCDTLGMADVGTVCDPKRSVIEDDG-LQAQATTAHELGHVFNPHDPAKHCASLNG 406
```

```

QY 383 KLRANHMSPSTLIQIDRANPWSACSAIITDLDGSHGDCILDDQSPKIPISLPEDLPQASY 442
Db 407 VTGDSHLMASLSDHSPWSPSCSAIYMTVSLDNGHGECLMDKPNQIKLPDLPGLY 466
QY 443 TLOQOCLAFYQSGKPCP-YMOYCTKLWCTGKAKGOMVOTRHFPAWDTSCGEGKCLK 501
Db 467 DANRQOFTFGESEKPCDAASTCTTLWCTGTSGLLVCOTKHFPAWDTSCGEGKWCVS 526
QY 502 GACVERHNLNKH---RVDSWAKWDPYGPCSRTCCGGVOLARRQCTNTPPANGKYCEGV 558
Db 527 GKCVNKTDM-KHFATPVHSGWPGWPGDCSRTCCGGVQYTMRECDNPVKNKGKCEGK 585
QY 559 RVYRSCNLEPCPSSAGSFRFEEQCEAFNGYNHSTNRLTLAVAKVPKYSVSPEDCKL 618
Db 586 RVYRSCNIEDCPDN-NGKTFREEQCEAHNEFSKASFGNEPTVETPKYAGVSPKDRCKL 644
QY 619 ICRANGTYFYVLAPKVDGTGLCSPDSTSVCGQKICAKGCDNGLGSKKRFDKCGVCGGD 678
Db 645 TCEAKGIGYFVLQPKVVDGTGCPSPDSTSVCGQCVKAGCDRIIDSKKFKDKCGVCGGN 704
QY 679 NKCKKVTGLFTKPMHGVFVVAIPAGASSIDIRORYKGLIGDDNYLALNKSQKYLIN 738
Db 705 GSTCKRMSGIVSTRGCHDIVTIPAGATNIEVKRNQKSRNNGSFLAIRAADGTIILN 764
QY 739 GHFVSAVERDLVKGSLRYSGTGPVAVESLQASRPILFVEVLSYKMTPPRVRYSF 798
Db 765 GNETSLTEQDLTYKGTVLRYSGSSAALERIRSFPLKEPLTIQVLMVGHALRPKIKFTY 824
QY 799 YLPEKREDKSSHPKDPGRPSVLHNSVLISNQVEQDDRPAPRVAGSWGSPCSASCSSG 858
Db 825 FMKKKTES-----FNAIPTFS-----EWIEWEGCSKTCSSG 857
QY 859 LQKRAVDCRGSAGQRTVPACDAAH--RPVETQACGE-PCPTW 897
Db 858 WQRVVOQRDINGH---PASECAKEVKPASTRCPADLPCHW 896

RESULT 15
AAW78435
ID AAW78435 standard; Protein; 727 AA.
XX
AC AAW78435:
XX
DT 11-MAY-1999 (first entry)
XX
DE Human ADAWTS-1 protein.
XX
KW Metalloproteinase-disintegrin protein with a thrombospondin domain; ARDS;
KW ADAWTS-1; drug composition; foodstuff; leukocyte; thrombocyte; hepatitis;
KW blood; erythrocyte; inflammatory disease; rheumatoid arthritis; asthma;
KW nephritis; Crohn's disease; acute respiratory disease syndrome.
XX
OS Homo sapiens.
XX
XX W09855643-A1.
PN
XX
PD 10-DEC-1998.
XX
XX 03-JUN-1998; 98WO-JP02449.
XX
XX 03-JUN-1997; 97JP-0160422.
XX
XX (KURE ) KUREHA CHEM IND CO LTD.
XX
XX Hakozaki M, Hirose K, Inoguchi E, Ishida Y, Ishioka K;
XX PI Kuno K, Matsushina K;
XX PI WPI; 1999-070277/06.
XX DR N-PSDB; AAX17990.
XX
XX Human metalloproteinase-disintegrin protein with thrombospondin
XX domain - useful as leukocyte and thrombocyte decreasing and
XX erythrocyte increasing agent

```

```

XX
PS
XX
CC This sequence represents a novel human metalloproteinase-disintegrin
CC protein with a thrombospondin domain (ADAWTS-1). The protein may be used
CC in drug compositions and foodstuffs, as an agent for decreasing the
CC leukocyte and thrombocyte blood count and increasing the erythrocyte
CC blood count, e.g. for treatment of inflammatory diseases such as
CC rheumatoid arthritis, hepatitis, nephritis, Crohn's disease, asthma
CC and ARDS.
XX
SQ Sequence 727 AA;
Query Match 44.0%; Score 2274; DB 20; Length 727;
Best Local Similarity 54.4%; Pred. No. 5.4e-168; Indels 40; Gaps 11;
Matches 408; Conservative 129; Mismatches 173;
QY 209 RAKFEFYSIPRYVETLVVADESVMVKFPGADLEHYLLTLATARLYRHPSILNPINVVVK 268
Db 9 RKREFVSSPRYVETLVVADQSNAEFPGSLKHLLTLFSAARLYKHPISIRNSVSLVVVK 68
QY 269 VLLLRDSDGPKVTGNAALTLNFCWAKKLNKVSOKHPEYWDTAILETRQDLGATGTC 328
Db 69 ILVIHDEQKPEVTSNAALTLNFCWAKKLNKVSOKHPEYWDTAILETRQDLGATGTC 128
QY 329 TLGNADVGTMCDPKRCSVIEDDGLPSAFTTAHELGHVFNPMHNVKVCYEEVFGKLRANH 388
Db 129 TLGNADVGTVCDPSRCSVIEDDGLQAFTTAHELGHVFNPMHNVKVCYEEVFGKLRANH 188
QY 389 MMSPTLIQIDRANPWSACSAIITDLDGSHGDCILDDQSPKIPISLPEDLPQASY 448
Db 189 MMSPTLIQIDRANPWSACSAIITDLDGSHGDCILDDQSPKIPISLPEDLPQASY 248
QY 449 ELAFYGVSKPCP-YMOYCTKLWCTGKAKGOMVOTRHFPAWDTSCGEGKCLKAGACVER 507
Db 249 QFTFGEDSKHCPDAASTCTLWCTGTSGVLVQCTKHFPAWDTSCGEGKCLKAGACVER 308
QY 508 HNLNKH---RVDSWAKWDPYGPCSRTCCGGVOLARRQCTNTPPANGKYCEGVRYKYS 564
Db 309 TD-RKHFDTPFHSGWPGWPGDCSRTCCGGVQYTMRECDNPVKNKGKCEGKRVYRS 367
QY 565 CNLEPCPSSAGSFRFEEQCEAFNGYNHSTNRLTLAVAWVPKYSVSPRDKCLICRANG 624
Db 368 CNLEDCPDN-NGKTFREEQCEAHNEFSKASFGSPGPAVEWIPKYAGVSPKDRCKLICQAKG 426
QY 625 TGYFYVLAPKVDGTGLCSPDSTSVCGQKICAKGCDNGLGSKKRFDKCGVCGGDNKSKK 684
Db 427 IGYFYVLQPKVVDGTGCPSPDSTSVCGQCVKAGCDRIIDSKKFKDKCGVCGGNGSTCKK 486
QY 685 VTGLFTKPMHGVFVVAIPAGASSIDIRORYKGLIGDDNYLALNKSQKYLINAGHFVYS 744
Db 487 ISGSVTSAPGYHDIVTIPTGATNIEVKORNQSRNNGSFLAIRAADGTIILNAGDYTL 546
QY 745 AVERDLVVGSLRYSGTGPVAVESLQASRPILFVEVLSYKMTPPRVRYSFYLPREP 804
Db 547 TLEQDINRYGVVLRYSGSSAALERIRSFPLKEPLTIQVLMVGHALRPKIKFTYFVKKK- 605
QY 805 REDKSSHPKDPGRPSVLHNSVLISNQVEQDDRPAPRVAGSWGSPCSASCGLQKRAV 864
Db 606 ---KES-----FNAIPTFS-----AWIEWEGCSKSCELGWQRRLV 639
QY 865 DCRGSAGQRTVPACDAAH--RPVETQACGE-PCPTWELSAWSPCSKSGRQFQRSLKCV 921
Db 640 ECRDINGQ---PASECAKEVKPASTRCPADHPCPQWLGEWSSCKTCKGKYKRSKCL 696
QY 922 GHGRLRLARDQCNLHRKQOE-LDFCVLRPC 950
Db 697 SHDGGVLSHESCDPLKKPKHFDICTLTQC 726

```

Search completed: April 29, 2003, 17:20:27
Job time : 61.2301 secs


```
|||||
Db 140 PNASPAQRNSQAHLLQRRGVPGSPGDPSTSRGCVASGWNFAILLRALDYPKPRAGFG 199
Qy 200 ESRSSRRSGRAKRVSTIPRVETLVVADESWMKFGADLEHYLLTLATAARLYRHPSIL 259
Db 200 ESRSSRRSGRAKRVSTIPRVETLVVADESWMKFGADLEHYLLTLATAARLYRHPSIL 259
Qy 260 NPINIVVVKVLLLRDRSGPKVGTGNAALTLRNFCAWOKKLNKYSDKHPEYWDTAIILFTQ 319
Db 260 NPINIVVVKVLLLRDRSGPKVGTGNAALTLRNFCAWOKKLNKYSDKHPEYWDTAIILFTQ 319
Qy 320 DLGATTCDTLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVCEE 379
Db 320 DLGATTCDTLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVCEE 379
Qy 380 VFGKLRANHMSPPTLIQIDRANPWSACSAIITDFLDSHGDCDLLDQPSKPISLPDLPG 439
Db 380 VFGKLRANHMSPPTLIQIDRANPWSACSAIITDFLDSHGDCDLLDQPSKPISLPDLPG 439
Qy 440 ASYTLSSQCELAFGVSGKPCPYMOYCTKLWCTGKAGQOMVQOTRHPFPWADGTSCGEGKLC 499
Db 440 ASYTLSSQCELAFGVSGKPCPYMOYCTKLWCTGKAGQOMVQOTRHPFPWADGTSCGEGKLC 499
Qy 500 LKGACVERHNLNKHVRDGSWAKDPYGPCSRTCGGGVQLARRQCTNPTPANGGKYCEGVR 559
Db 500 LKGACVERHNLNKHVRDGSWAKDPYGPCSRTCGGGVQLARRQCTNPTPANGGKYCEGVR 559
Qy 560 VKYRSCNLEPCPSSASGKSFRBOCEAFNGYNHSTNRLTLAVAVWPKYSGVSPRDCKLI 619
Db 560 VKYRSCNLEPCPSSASGKSFRBOCEAFNGYNHSTNRLTLAVAVWPKYSGVSPRDCKLI 619
Qy 620 CRANGTGYFYVLAPKVVDGTLCSFSDSTSVCVQKCIKAGCDGNLGSKKRDKCGVCGGDN 679
Db 620 CRANGTGYFYVLAPKVVDGTLCSFSDSTSVCVQKCIKAGCDGNLGSKKRDKCGVCGGDN 679
Qy 680 KSKKKVTGLFTKPMHGVNFVVAIPAGASSIDIRQGVKGLIGDDNYLALNKSQGYLLNG 739
Db 680 KSKKKVTGLFTKPMHGVNFVVAIPAGASSIDIRQGVKGLIGDDNYLALNKSQGYLLNG 739
Qy 740 HFVYSAVERDLVKGSLRLYSGTGTAVESLQASRPILPTVLSVKGKMTPPRVRYSFY 799
Db 740 HFVYSAVERDLVKGSLRLYSGTGTAVESLQASRPILPTVLSVKGKMTPPRVRYSFY 799
Qy 800 LPKEPREDKSHPKDRGSPVHLNSVLNSLNOVEQPDPRPARWAGSWGPCASCGSGL 859
Db 800 LPKEPREDKSHPKDRGSPVHLNSVLNSLNOVEQPDPRPARWAGSWGPCASCGSGL 859
Qy 860 QKRAVDCRGSAQRTVPACDAARHPVETQACGEPCTWELSAWSPCSKSCGRGFORSLK 919
Db 860 QKRAVDCRGSAQRTVPACDAARHPVETQACGEPCTWELSAWSPCSKSCGRGFORSLK 919
Qy 920 CVHGGRLRLARDQCNLHRKPOELDFCVLRPC 950
Db 920 CVHGGRLRLARDQCNLHRKPOELDFCVLRPC 950

RESULT 2
Q91Z56 PRELIMINARY; PRT; 340 AA.
AC Q91Z56;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE Similar to a disintegrin and metalloproteinase with thrombospondin
DE motifs 1 (ADAMTS-1) (Fragment).
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

|||||
Db 140 PNASPAQRNSQAHLLQRRGVPGSPGDPSTSRGCVASGWNFAILLRALDYPKPRAGFG 199
Qy 200 ESRSSRRSGRAKRVSTIPRVETLVVADESWMKFGADLEHYLLTLATAARLYRHPSIL 259
Db 200 ESRSSRRSGRAKRVSTIPRVETLVVADESWMKFGADLEHYLLTLATAARLYRHPSIL 259
Qy 260 NPINIVVVKVLLLRDRSGPKVGTGNAALTLRNFCAWOKKLNKYSDKHPEYWDTAIILFTQ 319
Db 260 NPINIVVVKVLLLRDRSGPKVGTGNAALTLRNFCAWOKKLNKYSDKHPEYWDTAIILFTQ 319
Qy 320 DLGATTCDTLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVCEE 379
Db 320 DLGATTCDTLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVCEE 379
Qy 380 VFGKLRANHMSPPTLIQIDRANPWSACSAIITDFLDSHGDCDLLDQPSKPISLPDLPG 439
Db 380 VFGKLRANHMSPPTLIQIDRANPWSACSAIITDFLDSHGDCDLLDQPSKPISLPDLPG 439
Qy 440 ASYTLSSQCELAFGVSGKPCPYMOYCTKLWCTGKAGQOMVQOTRHPFPWADGTSCGEGKLC 499
Db 440 ASYTLSSQCELAFGVSGKPCPYMOYCTKLWCTGKAGQOMVQOTRHPFPWADGTSCGEGKLC 499
Qy 500 LKGACVERHNLNKHVRDGSWAKDPYGPCSRTCGGGVQLARRQCTNPTPANGGKYCEGVR 559
Db 500 LKGACVERHNLNKHVRDGSWAKDPYGPCSRTCGGGVQLARRQCTNPTPANGGKYCEGVR 559
Qy 560 VKYRSCNLEPCPSSASGKSFRBOCEAFNGYNHSTNRLTLAVAVWPKYSGVSPRDCKLI 619
Db 560 VKYRSCNLEPCPSSASGKSFRBOCEAFNGYNHSTNRLTLAVAVWPKYSGVSPRDCKLI 619
Qy 620 CRANGTGYFYVLAPKVVDGTLCSFSDSTSVCVQKCIKAGCDGNLGSKKRDKCGVCGGDN 679
Db 620 CRANGTGYFYVLAPKVVDGTLCSFSDSTSVCVQKCIKAGCDGNLGSKKRDKCGVCGGDN 679
Qy 680 KSKKKVTGLFTKPMHGVNFVVAIPAGASSIDIRQGVKGLIGDDNYLALNKSQGYLLNG 739
Db 680 KSKKKVTGLFTKPMHGVNFVVAIPAGASSIDIRQGVKGLIGDDNYLALNKSQGYLLNG 739
Qy 740 HFVYSAVERDLVKGSLRLYSGTGTAVESLQASRPILPTVLSVKGKMTPPRVRYSFY 799
Db 740 HFVYSAVERDLVKGSLRLYSGTGTAVESLQASRPILPTVLSVKGKMTPPRVRYSFY 799
Qy 800 LPKEPREDKSHPKDRGSPVHLNSVLNSLNOVEQPDPRPARWAGSWGPCASCGSGL 859
Db 800 LPKEPREDKSHPKDRGSPVHLNSVLNSLNOVEQPDPRPARWAGSWGPCASCGSGL 859
Qy 860 QKRAVDCRGSAQRTVPACDAARHPVETQACGEPCTWELSAWSPCSKSCGRGFORSLK 919
Db 860 QKRAVDCRGSAQRTVPACDAARHPVETQACGEPCTWELSAWSPCSKSCGRGFORSLK 919
Qy 920 CVHGGRLRLARDQCNLHRKPOELDFCVLRPC 950
Db 920 CVHGGRLRLARDQCNLHRKPOELDFCVLRPC 950

Query Match 1.1%; Score 10; DB 6; Length 192;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 360 AHELGHVFN 369
Db 29 AHELGHVFN 38
|||||

RESULT 4
Q8TE57 PRELIMINARY; PRT; 1072 AA.
AC Q8TE57;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Metalloprotease disintegrin 16 with thrombospondin type I motif.
GN ADAMTS16.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
```

```
DR EMBL: BC009667; AAH09667.1; -
DR InterPro: IP8000884; TSP1.
DR Pfam: PF00090; tsp_1; 2.
KW Integrin.
FT NON_TER 1
SQ SEQUENCE 340 AA; 37021 MW; 06562C747634A8BD CRC64;

Query Match 7.5%; Score 71; DB 11; Length 340;
Best Local Similarity 100.0%; Pred. No. 1.1e-67;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 685 VTGLETKPMHGVNFVVAIPAGASSIDIRQGVKGLIGDDNYLALNKSQGYLLNGHFVYS 744
Db 75 VTGLETKPMHGVNFVVAIPAGASSIDIRQGVKGLIGDDNYLALNKSQGYLLNGHFVYS 134

Qy 745 AVERDLVWVKS 755
Db 135 AVERDLVWVKS 145

RESULT 3
Q95N24 PRELIMINARY; PRT; 192 AA.
ID Q95N24;
AC Q95N24;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Aggrecaanase-1 (Fragment).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RA Flannery C.R., Little C.B.;
RT "Expression and activity of equine aggrecaanases.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF368321; AAK53425.1; -
DR MEROPS: M12.221; -
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000130; Zn_MTPeptide.
DR PROSITE: PS50215; ADAM_MEPRO; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 192
SQ SEQUENCE 192 AA; 20670 MW; 9013B0E19FCE8C56 CRC64;

Query Match 1.1%; Score 10; DB 6; Length 192;
Best Local Similarity 100.0%; Pred. No. 0.091;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL: AJ315734; CAC86015.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1072 AA; 119491 MW; 081EEFF78F47D061 CRC64;
Query Match 1.1%; Score 10; DB 4; Length 1072;
Best Local Similarity 100.0%; Pred. No. 0.43;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 527 PCSRTCGGV 536
Db 597 PCSRTCGGV 606
RESULT 5
Q8TEY8 PRELIMINARY; PRT; 1159 AA.
AC Q8TEY8
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ADAMTS14.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21839041; PubMed-11741898;
RA Collige A., Vandenbergh I., Thiry M., Lambert C.A., Van Beeumen J.,
RA Li S.W., Prockop D.J., Lapiere C.M., Nusgens B.V.;
RT "Cloning and Characterization of ADAMTS-14, a Novel ADAMTS Displaying
RT High Homology with ADAMTS-2 and ADAMTS-3";
RL J. Biol. Chem. 277:5756-5766(2002).
DR EMBL: AF366351; AAL79814.1; -.
SQ SEQUENCE 1159 AA; 127336 MW; A5B130149BF7FF34 CRC64;
Query Match 1.1%; Score 10; DB 4; Length 1159;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 670 DKCGVCGGDN 679
Db 637 DKCGVCGGDN 646
RESULT 6
Q8WXS8 PRELIMINARY; PRT; 1223 AA.
AC Q8WXS8
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE A disintegrin-like and metalloprotease with thrombospondin type 1
DE motif 14 precursor.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21638061; PubMed-11779638;
RA Bolz H., Ramirez A., von Brderlow B., Kubisch C.;
RT "Characterization of ADAMTS14, a novel member of the ADAMTS
RT metalloproteinase family";
RL Biochim. Biophys. Acta 1522:221-225(2001).
DR EMBL: AF358666; AAL40229.1; -.

DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSP1.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF00090; tsp_1; 4.
DR SMART: SM00209; TSP1; 4.
DR PROSITE: PS50215; ADAM_MEPRO; 1.
DR PROSITE: PS00092; TSP1; 1.
KW Signal; Integrin; Protease; Metalloprotease.
FT SIGNAL 1 22 POTENTIAL.
SQ SEQUENCE 1223 AA; 133871 MW; 36394AC4D92F170F CRC64;
Query Match 1.1%; Score 10; DB 4; Length 1223;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 670 DKCGVCGGDN 679
Db 701 DKCGVCGGDN 710
RESULT 7
Q8TE55 PRELIMINARY; PRT; 1223 AA.
AC Q8TE55
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Metalloprotease-disintegrin protease.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21856482; PubMed-11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL: AJ345098; CAC87943.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1223 AA; 133925 MW; D585B6593977ED15 CRC64;
Query Match 1.1%; Score 10; DB 4; Length 1223;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 670 DKCGVCGGDN 679
Db 701 DKCGVCGGDN 710
RESULT 8
Q95428 PRELIMINARY; PRT; 1235 AA.
AC Q95428
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical 133.5 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Madan A., Qin S., Abbasi N., Dors M., Ratcliffe A.,

RA Madan A., Dickhoff R., Shaffer T., James R., Lasky S., Hood L.;
RT "Complete sequence of the gene for presenilin 1";
RL EMBL; AF109907; AAC97963.1; -;
DR HSSP; P12111; 2KNT.
DR InterPro; IPR003598; Iq_c2.
DR InterPro; IPR003006; Iq_MHC.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00047; Iq; 3.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00090; tsp_1; 5.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 1.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00131; KU; 1.
DR SMART; SM00209; TSPI; 5.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 2.
DR PROSITE; PS00092; TSPI; 4.
KW Hypothetical protein; Immunoglobulin domain;
KW Serine protease inhibitor.
SQ SEQUENCE 1235 AA; 133477 MW; A0B44CCE4F38E350 CRC64;

Query Match 1.1%; Score 10; DB 4; Length 1235;
Best Local Similarity 100.0%; Pred. No. 0.49;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 527 PCSTCGGV 536
DB 37 PCSTCGGV 46
|||||

RESULT 9
O44938 PRELIMINARY; PRT; 1572 AA.
ID O44938;
AC O44938;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Thrombospondin.
GN Thrombospondin.
OS Haemochus contortus (Barber pole worm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Trichostrongyloidea; Haemonchidae; Haemonchinae; Haemonchus.
OX NCBI_TaxID=6289;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOREDUN;
RA Skuce P.J., Newlands G.F.J., Stewart M., Pettit D., Smith D.,
RT Knox D.P.;
RT "Cloning and characterization of thrombospondin, a novel multidomain
RT glycoprotein associated with the gut of Haemonchus contortus";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF043121; AAB99830.2; -;
DR HSSP; P05067; ICAO.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF00014; Kunitz_BPTI; 6.
DR Pfam; PF00090; tsp_1; 6.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 6.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00131; KU; 6.
DR SMART; SM00209; TSPI; 7.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 6.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 6.
DR PROSITE; PS00092; TSPI; 2.
KW Serine protease inhibitor.
SQ SEQUENCE 1572 AA; 171871 MW; 2260B30DC2F903EC CRC64;

Query Match 1.1%; Score 10; DB 5; Length 1572;

Best Local Similarity 100.0%; Pred. No. 0.61;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTCGGV 537
DB 83 CSRTCGGV 92
|||||

RESULT 10
Q19791 PRELIMINARY; PRT; 2165 AA.
ID Q19791; Q27524;
AC Q19791; Q27524;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE F25H8.3 protein.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Gajadaty S.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Smailson M., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RL Nature 368:32-38(1994).
DR EMBL; 269361; CAA93288.1; -;
DR EMBL; 269360; CAA93288.1; JOINED.
DR EMBL; 269360; CAA93287.1; -;
DR EMBL; 269361; CAA93287.1; JOINED.
DR HSSP; P15167; 1DTH.
DR MEROPS; M12.135; -;
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR00130; Zn_MTpeptidse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 14.
DR SMART; SM00209; TSPI; 18.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSPI; 6.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 2165 AA; 244397 MW; FCC3DA8AAA9C4888 CRC64;

Query Match 1.1%; Score 10; DB 5; Length 2165;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTCGGV 537
DB 615 CSRTCGGV 624
|||||

RESULT 11
Q908G8 PRELIMINARY; PRT; 3198 AA.
ID Q908G8
AC Q908G8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE lacunin precursor.
OS Manduca sexta (Tobacco hawkmoth) (Tracheata; Hexapoda; Insecta).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Spingioidea; Spingidae; Spinginae; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99457716; PubMed=10528409;
RX Nardi J.B., Martos R., Walden K.K., Lampe D.J., Robertson H.M.;
RT "Expression of lacunin, a large multidomain extracellular matrix
protein, accompanies morphogenesis of epithelial monolayers in Manduca
sexta.";
RL Insect Biochem. Mol. Biol. 29:883-897(1999).
DR EMBL; AF078161; AAF04457.1; -.
DR HSSP; P12111; 2KNT.
DR InterPro; IPR004094; Antistatin.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR002221; WAP.
DR Pfam; PF02822; Antistatin; 4.
DR Pfam; PF00047; Ig; 2.
DR Pfam; PF00014; Kunitz_BPTI; 10.
DR Pfam; PF00095; Wap; 1.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 10.
DR SMART; SM00408; IGC2; 2.
DR SMART; SM00131; KU; 10.
DR SMART; SM00209; TSPI; 7.
DR SMART; SM00217; WAP; 1.
DR PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 8.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 10.
DR PROSITE; PS00092; TSPI; 1.
KW Immunoglobulin domain; Serine protease inhibitor; Signal.
FT SIGNAL 1 21 POTENTIAL.
SQ SEQUENCE 3198 AA; 349364 MW; AB4ACD459CD09134 CRC64;

Query Match 1.18; Score 10; DB 5; Length 3198;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 527 PCSRTCGGV 536
DB 71 PCSRTCGGV 80
|||||

RESULT 12
Q9GL54
ID Q9GL54 PRELIMINARY; PRT; 269 AA.
AC Q9GL54;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Aggrecanase-2 (Fragment).
GN ADAMTS-11.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Goad D.L., Goad M.E.;
RT "Molecular cloning of lapine aggrecanase-2 (ADAMTS-11) from articular
chondrocytes.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF317415; AAG33062.1; -.
DR HSSP; O9PW35; 1BUD.
DR MEROPS; M12.225; -.
DR InterPro; IPR001590; Reprolysin.

DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_MTpeptidse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS02115; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSPI; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1 1
FT NON_TER 269 269
SQ SEQUENCE 269 AA; 29193 MW; 97A1CA80B33452FA CRC64;

Query Match 0.9%; Score 9; DB 6; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 TLGMADVGT 337
DB 1 TLGMADVGT 9
|||||

RESULT 13
Q8TE60
ID Q8TE60 PRELIMINARY; PRT; 1081 AA.
AC Q8TE60;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE ADAMTS18 protein.
GN ADAMTS18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
seven novel human ADAMTS, a family of metalloproteinases with
disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ311903; CAC83612.1; -.
SQ SEQUENCE 1081 AA; 119655 MW; 0438BF645676461E CRC64;

Query Match 0.9%; Score 9; DB 4; Length 1081;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTCGGV 536
DB 600 CSRTCGGV 608
|||||

RESULT 14
Q9EPX2
ID Q9EPX2 PRELIMINARY; PRT; 1280 AA.
AC Q9EPX2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Papilin.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1;
RX MEDLINE=20530499; PubMed=11076767;
RA Kramerova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,
RA Kramerov A.A., Kusche-Gullberg M., Kramer J.M., Ackley B.D.,

```

RA Sieron A.L., Prockop D.J., Fessler J.H.;
RT "Papilin in development; a pericellular protein with a homology to the
RL ADAMTS metalloproteinases.";
RL Development 127:5475-5485(2000).
DR EMBL: AF314171; AAG41980.1; -.
DR HSSP: P12111; 2KNT.
DR InterPro: IPR000183; Decarboxylase2.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003600; Ig_like.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR000884; TSPI.
DR Pfam: PF00047; Ig; 3.
DR Pfam: PF00014; Kunitz_BPTI; 1.
DR PRINTS: PR00759; BASICPTASE.
DR PRODOM: PD000222; Kunitz_BPTI; 1.
DR SMART: SM00408; Igc2; 3.
DR SMART: SM00409; Igc2; 2.
DR SMART: SM00131; KU; 3.
DR SMART: SM00209; TSPI; 7.
DR SMART: SM00217; WAP; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE: PS00279; BPTI_KUNITZ_2; 1.
DR PROSITE: PS00879; ODR_DC_2_2; UNKNOWN_1.
DR PROSITE: PS50092; TSPI; 3.
KW Immunoglobulin domain; Serine protease inhibitor.
SQ SEQUENCE 1280 AA; 138824 MW; AE287705E561AF30 CRC64;

Query Match 0.9%; Score 9; DB 1; Length 1280;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 527 CSRTCGGG 535
DB 38 CSRTCGGG 46

RESULT 15
Q9GORO PRELIMINARY; PRT; 2174 AA.
AC Q9GORO;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Extracellular matrix protein papilin precursor.
GN PPN OR CG1540 OR CG18436.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DP CN BW;
RX MEDLINE=20530499; PubMed=11076767;
RA Kramerova I.A., Kawaguchi N., Nelson R.E., Fessler L.I., Chen Y.,
RA Kramerov A.A., Kusche-Gullberg M., Kramer J.H.;
RA Sieron A.L., Prockop D.J., Fessler J.H.;
RT "Papilin in development; a pericellular protein with a homology to the
RL ADAMTS metalloproteinases.";
RL Development 127:5475-5485(2000).
DR EMBL: AF205357; AAG37995.1; -.
DR HSSP: P12111; 2KNT.
DR FLYBase: FBgn003137; Ppn.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003600; Ig_like.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR002223; Kunitz_BPTI.
DR InterPro: IPR000884; TSPI.

```

```

DR InterPro: IPR002221; WAP.
DR Pfam: PF00047; Ig; 2.
DR Pfam: PF00014; Kunitz_BPTI; 3.
DR Pfam: PF00090; tsp_1; 5.
DR Pfam: PF00095; wap; 1.
DR PRINTS: PR00003; 4DISULPHCORE.
DR PRODOM: PD000222; Kunitz_BPTI; 3.
DR SMART: SM00409; Ig; 3.
DR SMART: SM00408; Igc2; 2.
DR SMART: SM00410; IG_like; 1.
DR SMART: SM00131; KU; 3.
DR SMART: SM00209; TSPI; 7.
DR SMART: SM00217; WAP; 1.
DR PROSITE: PS00317; 4-DISULFIDE_CORE; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1; 3.
DR PROSITE: PS00279; BPTI_KUNITZ_2; 3.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS50092; TSPI; 3.
KW Matrix protein; Serine protease inhibitor; Signal.
FT SIGNAL 1 26
SQ SEQUENCE 2174 AA; 231936 MW; 038F707952623120 CRC64;

Query Match 0.9%; Score 9; DB 5; Length 2174;
Best Local Similarity 100.0%; Pred. No. 9.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTCGGGV 536
DB 69 CSRTCGGGV 77

```

Search completed: April 29, 2003, 17:23:43
Job time : 67 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:22:09 ; Search time 27 seconds
(without alignments)

3382.507 Million cell updates/sec

Title: US-10-009-332-1

Perfect score: 950

Sequence: 1 MLLLGILTAFAAGTAGGFE.....DQCNLHRKPOELDFCVLRPC 950

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR.73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17	1.8	951	2	T00017	gene ADAMTS-1 prot
2	14	1.5	550	2	T47158	hypothetical prote
3	13	1.4	837	2	T00355	hypothetical prote
4	10	1.1	1205	2	T18517	procollagen N-endo
5	10	1.1	2165	2	T21371	hypothetical prote
6	8	0.8	196	2	AF0715	probable ABC trans
7	8	0.8	206	2	A45517	coccidioidis-relate
8	8	0.8	407	2	S66260	metalloproteinase
9	8	0.8	440	2	D95029	PTS system, IIC co
10	8	0.8	440	2	G97900	hypothetical prote
11	8	0.8	607	2	T11032	NADH2 dehydrogenas
12	8	0.8	608	2	A81293	probable flagellar
13	8	0.8	712	2	A45638	immunodominant mic
14	8	0.8	788	2	T25061	hypothetical prote
15	8	0.8	1049	2	T30525	alpha-mannosidase
16	8	0.8	5037	1	A54161	ryanodine-binding
17	7	0.7	66	2	AB2371	hypothetical prote
18	7	0.7	96	2	I68623	neurofibromin - mo
19	7	0.7	108	2	A12333	hypothetical prote
20	7	0.7	110	2	C72472	hypothetical prote
21	7	0.7	111	2	B26567	nitrogen regulator
22	7	0.7	112	2	B13078	nitrogen regulator
23	7	0.7	112	2	S52328	nitrogen regulator
24	7	0.7	112	2	C87493	nitrogen regulator
25	7	0.7	112	2	D82985	nitrogen regulator
26	7	0.7	112	2	AD32794	nitrogen regulator
27	7	0.7	112	2	AD3374	nitrogen regulator
28	7	0.7	119	2	G81738	hypothetical prote
29	7	0.7	124	2	AH5226	conserved hypothet

30	7	0.7	125	2	A83048	hypothetical prote
31	7	0.7	131	1	D69827	hypothetical prote
32	7	0.7	132	2	A10244	probable rhodanese
33	7	0.7	135	1	B49205	virulence-associat
34	7	0.7	138	2	B70867	hypothetical prote
35	7	0.7	142	2	C70059	hypothetical prote
36	7	0.7	143	2	E64359	ribosomal protein
37	7	0.7	146	2	S51810	trypsin inhibitor
38	7	0.7	157	2	A97573	nitrogen regulator
39	7	0.7	160	2	G71122	hypothetical prote
40	7	0.7	161	2	T11855	DnaJ protein homol
41	7	0.7	163	2	B70559	hypothetical prote
42	7	0.7	167	2	E87476	conserved hypothet
43	7	0.7	172	2	JH0780	trypsin inhibitor
44	7	0.7	172	2	A24082	trypsin inhibitor
45	7	0.7	172	2	A27220	trypsin inhibitor

ALIGNMENTS

RESULT 1

T00017

gene ADAMTS-1 protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000

C:Accession: T00017

R:Kuno, K.; Lizasa, H.; Ohno, S.; Matsushima, K.

Genomics 46, 466-471, 1997

A:Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1

A:Reference number: Z14055; MUID:98110583; PMID:9441751

A:Accession: T00017

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-951 <NUN>

A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BA24501.1; PID:g2809057

A:Experimental source: strain 129SVJ

C:Genetics:

A:Gene: ADAMTS-1

A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 603/1; 660/3; 719/2

C:Superfamily: thrombospondin type 1 repeat homology

F:542-598/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 1.8%; Score 17; DB 2; Length 951;

Best Local Similarity 100.0%; Pred. No. 1.9e-08;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNPHD 372

|||||

Db 380 AFTTAHELGHVFNPHD 396

RESULT 2

T47158

hypothetical protein DKFZp762C1110.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47158

R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.

submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24379

A:Accession: T47158

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-550 <AAA>

A:Cross-references: EMBL:AL162080

A:Experimental source: adult melanoma (MeWo cell line); clone DKFZp762C1110

C:Genetics:

A:Note: DKFZp762C1110.1

Query Match

Best Local Similarity 100.0%; Pred. No. 1.5e-05;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 484 HFPWADTSGEGK 497
|||||
Db 108 HFPWADTSGEGK 121

RESULT 3

T00355
hypothetical protein KIAA0688 - human
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00355
R: Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5: 169-176, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A:Reference number: Z14142; MUID:98403880; PMID:9734811
A:Accession: T00355
A:Status: preliminary;
A:Molecule type: mRNA
A:Residues: 1-837 <ISH>
A:Cross-references: EMBL:AB014588; NID:g3327189; PIDN:BAA31663.1; PID:g3327190
A:Experimental source: Brain
C:Genetics:
A:Gene: KIAA0688
C:Superfamily: thrombospondin type 1 repeat homology
F:519-575/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 1.48; Score 13; DB 2; Length 837;
Best Local Similarity 100.0%; Pred. No. 0.0023; Indels 0; Gaps 0;
Matches 13; Conservative 0; Mismatches 0;

QY 311 DTAILFTQDLCG 323
|||||
Db 311 DTAILFTQDLCG 323

RESULT 4

T18517
collagen N-endopeptidase (EC 3.4.24.14) I - bovine
N:Alternate names: procollagen N-proteinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18517
R: Colige, A.; Nussgens, B.V.; Lapiere, C.M.
submitted to the EMBL Data Library, February 1996
A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.
A:Reference number: Z18941
A:Accession: T18517
A:Status: preliminary;
A:Molecule type: mRNA
A:Residues: 1-1205 <COL>
A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1
A:Experimental source: skin
C:Genetics:
A:Gene: PC I-NP
C:Function:
A:Description: catalyzes cleavage of the propeptides of type I and II collagens prior to
C:Keywords: hydrolase; metalloproteinase

Query Match 1.1%; Score 10; DB 2; Length 1205;
Best Local Similarity 100.0%; Pred. No. 0.39;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 DRKGVCGGDN 679
|||||
Db 704 DRKGVCGGDN 713

RESULT 5

T21371
hypothetical protein F25H8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T21371; T24896
R: Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19413
A:Accession: T21371
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WIL>
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone F25H8
R: Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19949
A:Accession: T24896
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WIL>
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone T13H10
C:Genetics:
A:Gene: CESP:F25H8.3
A:Map position: 4
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1;

Query Match 1.1%; Score 10; DB 2; Length 2165;
Best Local Similarity 100.0%; Pred. No. 0.63;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTCGG3VQ 537
|||||
Db 615 CSRTCGG3VQ 624

RESULT 6

AF0715
probable ABC transport ATP-binding chain STY1861 [imported] - Salmonella enterica sub
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AF0715
R: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
S.; Moulie, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: AF0715
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-196 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD02095.1; PID:g16502930; GSPDB:GN00176
C:Genetics:
A:Gene: STY1861

Query Match 0.8%; Score 8; DB 2; Length 196;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 AIRRALDP 190
|||||
Db 131 AIRRALDP 138

RESULT 7

A45517
cocciidiosis-related antigen - Eimeria tenella (fragment)
N:Alternate names: thrombospondin-related antigen, 100K
C:Species: Eimeria tenella
C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 07-May-1999
C:Accession: A45517; S27818
R: Clarke, L.E.; Tomley, F.M.; Wisner, M.H.; Foulds, I.J.; Boursnell, M.E.G.
Mol. Biochem. Parasitol. 41, 269-280, 1990

A:Title: Regions of an Eimeria tenella antigen contain sequences which are conserved in
A:Reference number: A45517; MUID:90377296; PMID:2204833
A:Accession: A45517
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <CIA>
A:Cross-references: GB:M32988; NID:gl58870; PID:gl58871
C:Superfamily: thrombospondin type 1 repeat homology
F:136-198/Domain: thrombospondin type 1 repeat homology <THR5>

Query Match 0.8%; Score 8; DB 2; Length 206;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 528 CSRTC GGG 535
|||||
DB 88 CSRTC GGG 95

RESULT 8
S66260
A:Title: Metalloproteinase (EC 3.4.24.-) precursor - southern copperhead
C:Species: Agkistrodon contortrix contortrix (southern copperhead)
C:Date: 17-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 22-Jun-1999
C:Accession: S66260; S74263
R:Selistre de Araujo, H.S.; Ownby, C.L.
Arch. Biochem. Biophys. 320, 141-148, 1995
A:Title: Molecular cloning and sequence analysis of cDNAs for metalloproteinases from br
A:Reference number: S66259; MUID:95314311; PMID:7793974
A:Accession: S66260
A:Molecule type: mRNA
A:Residues: 1-407 <SEL>
A:Cross-references: EMBL:U18234; NID:g603216; PIDN:AA59704.1; PID:g603217
A:Accession: S74263
A:Molecule type: protein
A:Residues: 188-206 <SEA>
C:Superfamily: atrolysin C
C:Keywords: hydrolase; metalloproteinase; zinc
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-187/Domain: signal sequence #status predicted <PRO>
F:188-407/Product: metalloproteinase #status experimental <WAT>
F:304-384,344-351,346-358/Disulfide bonds: #status predicted
F:329,333,339/Binding site: zinc, catalytic (His) #status predicted
F:330/Active site: Glu #status predicted

Query Match 0.8%; Score 8; DB 2; Length 407;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 335 VGTMC DPK 342
|||||
DB 300 VGTMC DPK 307

RESULT 9
D95029
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: D95029
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-440 <KUR>
A:Cross-references: GB:AB005672; PIDN:AAK74429.1; PID:gl4971719; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4

PTB system, IIC component [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: D95029
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: D95029
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-440 <KUR>
A:Cross-references: GB:AB005672; PIDN:AAK74429.1; PID:gl4971719; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4

C:Genetics:
A:Gene: SPO250
C:Superfamily: phosphotransferase system enzyme II factor II, phosphoenolpyruvate-de

Query Match 0.8%; Score 8; DB 2; Length 440;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 VPLOGLTG 93
|||||
DB 226 VPLOGLTG 233

RESULT 10
G97900
A:Title: hypothetical protein PTS-EIIC [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: G97900
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: G97900
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-440 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99035.1; PID:gl5457778; GSPDB:GN00174
C:Genetics:
A:Gene: PTS-EIIC
C:Superfamily: phosphotransferase system enzyme II factor II, phosphoenolpyruvate-de

Query Match 0.8%; Score 8; DB 2; Length 440;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 VPLOGLTG 93
|||||
DB 226 VPLOGLTG 233

RESULT 11
T11032
A:Title: dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - redhead mitochondrion
C:Species: mitochondrion Aythya americana (redhead)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002
C:Accession: T11032
R:Mindell, D.P.; Sorenson, M.D.; Dimcheff, D.E.
Proc. Natl. Acad. Sci. U.S.A. 95, 10693-10697, 1998
A:Title: Multiple independent origins of mitochondrial gene order in birds.
A:Reference number: 217242
A:Accession: T11032
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-607 <JOH>
A:Cross-references: EMBL:AF090337; NID:g4887659; PID:g4887670; PIDN:AA32262.1
C:Genetics:
A:Gene: mitochondrion
A:Genetic code: SGI1
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylati

Query Match 0.8%; Score 8; DB 2; Length 607;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 242 LTLTLATA 249
|||||
DB 411 LTLTLATA 418

RESULT 12

A81293
probable flagellar hook-associated protein Cj1466 [imported] - Campylobacter jejuni (str
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: A81293
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: A81293
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-608 <PAR>
A:CROSS-references: GB:AL139078; GB:AL111168; NID:g6968723; PIDN:CAB73889.1; PID:g696889
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: flgK; Cj1466

Query Match 0.8%; Score 8; DB 2; Length 608;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 GTGTAVES 768

Db 57 GTGTAVES 64

RESULT 13

A45638
immunodominant microneme protein Etp100 - Eimeria tenella
C:Species: Eimeria tenella
C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A45638
R:Tomley, F.M.; Clarke, L.E.; Kawazoe, U.; DiJkema, R.; Kok, J.J.
Mol. Biochem. Parasitol. 49, 277-288, 1991
A:Title: Sequence of the gene encoding an immunodominant microneme protein of Eimeria te
A:Reference number: A45638; MUID:92131064; PMID:1775171
A:Accession: A45638
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-712 <TOM>
A:CROSS-references: GB:AF032905; GB:M73495; NID:g2707732; PIDN:AAD03350.1; PID:g2707733
A:Note: sequence extracted from NCBI backbone (NCBIN:77752, NCBI:77756)
C:Superfamily: thrombospondin type 1 repeat homology; von Willebrand factor type A repea
F:48-218/Domain: von Willebrand factor type A repeat homology <VWAL>
F:238-296/Domain: thrombospondin type 1 repeat homology <THR1>
F:309-371/Domain: thrombospondin type 1 repeat homology <THR2>
F:372-432/Domain: thrombospondin type 1 repeat homology <THR3>
F:433-493/Domain: thrombospondin type 1 repeat homology <THR4>
F:494-556/Domain: thrombospondin type 1 repeat homology <THR5>
F:560-610/Domain: thrombospondin type 1 repeat homology <THR6>

Query Match 0.8%; Score 8; DB 2; Length 712;

Best Local Similarity 100.0%; Pred. No. 29;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTC GGG 535

Db 446 CSRTC GGG 453

RESULT 14

T25061
hypothetical protein T21B6.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T25061
R:Cottage, A.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19975

A:Accession: T25061

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-788 <WIL>

A:CROSS-references: EMBL:Z68011; PIDN:CAA92014.1; GSPDB:GN00028; CESP:T21B6.3

A:Experimental source: clone T21B6

C:Genetics:

A:Gene: CESP:T21B6.3

A:Map position: X

A:Introns: 20/1; 47/1; 76/1; 152/1; 735/2; 754/2

Query Match 0.8%; Score 8; DB 2; Length 788;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 851 CSASCGSG 858

Db 514 CSASCGSG 521

RESULT 15

T30525

alpha-mannosidase (EC 3.2.1.24) - Emericella nidulans

C:Species: Emericella nidulans, Aspergillus nidulans

C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000

C:Accession: T30525

R:Eades, C.J.; Gilbert, A.; Goodman, C.D.; Hintz, W.E.

Glycobiology 8, 17-33, 1998

A:Title: Identification and analysis of a class 2 alpha-mannosidase in Aspergillus ni

A:Reference number: Z20843; MUID:98119762; PMID:9451011

A:Accession: T30525

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1049 <EAD>

A:CROSS-references: EMBL:AF016850; NID:g2407175; PID:g2407176; PIDN:AAB70514.1

C:Genetics:

A:Note: msd2

A:Introns: 45/2; 219/1; 877/2

C:Superfamily: Saccharomyces alpha-mannosidase

C:Keywords: glycosidase; hydrolase

Query Match 0.8%; Score 8; DB 2; Length 1049;

Best Local Similarity 100.0%; Pred. No. 40;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 PLQGLTGG 94

Db 138 PLQGLTGG 145

Search completed: April 29, 2003, 17:25:03

Job time : 32 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:18:08 ; Search time 16 Seconds
(without alignments)

2462.657 Million cell updates/sec

Title: US-10-009-332-1

Perfect score: 950

Sequence: 1 MLLGLTLTAFAGRTAGGFE.....DQCNLHRKPOELDFCVLRPC 950

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	1.8	967	1	ATSL_HUMAN
2	17	1.8	967	1	ATSL_RAT
3	17	1.8	968	1	ATSL_MOUSE
4	13	1.4	630	1	ATSL_MOUSE
5	13	1.4	837	1	ATSL_HUMAN
6	13	1.4	1629	1	ATSL_HUMAN
7	12	1.3	905	1	ATSL_MOUSE
8	11	1.2	930	1	ATSL_HUMAN
9	11	1.2	930	1	ATSL_MOUSE
10	10	1.1	890	1	ATSL_HUMAN
11	10	1.1	1077	1	ATSL_HUMAN
12	10	1.1	1205	1	ATSL_BOVIN
13	10	1.1	1205	1	ATSL_HUMAN
14	10	1.1	1211	1	ATSL_HUMAN
15	9	0.9	860	1	ATSL_HUMAN
16	8	0.8	245	1	ATSL_BOVIN
17	8	0.8	408	1	IHH_CHICK
18	8	0.8	790	1	AD30_HUMAN
19	8	0.8	886	1	SM6B_MOUSE
20	8	0.8	887	1	SM6B_RAT
21	8	0.8	888	1	SM6B_HUMAN
22	8	0.8	997	1	ATSL_HUMAN
23	8	0.8	1593	1	ATSL_HUMAN
24	7	0.7	111	1	GLNB_RHIL
25	7	0.7	112	1	GLNB_AZOB
26	7	0.7	112	1	GLNB_RHIL
27	7	0.7	112	1	GLNB_RHIL
28	7	0.7	112	1	GLNB_RHIL
29	7	0.7	112	1	GLNB_RHIL
30	7	0.7	121	1	IAX1_WHEAT
31	7	0.7	131	1	CRB2_BACSU
32	7	0.7	143	1	RL15_METJA
33	7	0.7	172	1	IDES_ERYLA

34 7 0.7 172 1 IDE3_ERYLA
35 7 0.7 172 1 IDE3_ERYLA
36 7 0.7 176 1 CFTR_MACMU
37 7 0.7 199 1 SP25_DROME
38 7 0.7 228 1 VG77_HSV1
39 7 0.7 244 1 FRDB_PROVU
40 7 0.7 253 1 YE24_MYCTU
41 7 0.7 274 1 PKI_NPVOP
42 7 0.7 302 1 CITG_KLEPN
43 7 0.7 311 1 PRMA_BACSU
44 7 0.7 317 1 GGH_RAT
45 7 0.7 321 1 K6PF_AQUAE

ALIGNMENTS

RESULT 1
ATSL_HUMAN STANDARD: PRT: 967 AA.
ID ATSL_HUMAN Q9UH83; Q9P2K0; Q9NSJ8;
AC Q9UH83; Q9UP80; Q9UH83; Q9P2K0; Q9NSJ8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).
GN ADAMTS1 OR METH1 OR KIAA1346.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Casas C., Pritchard M.A., Estivill X., Arbones M.L.;
RT "Cloning, characterization and mapping on human chromosome 21 of the
RT orthologue of murine Adamts-1";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC TISSUE=Heart;
RX MEDLINE=99367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
RA Lombardo M., Iruela-Arispe M.L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
RT family of proteins with angio-inhibitory activity.";
RL J. Biol. Chem. 274:23349-23357(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Endothelial cells;
RX MEDLINE=20247184; PubMed=10785405;
RA Glienke J., Schmitt A.O., Pillarsky C., Hinzmann B., Weiss B.,
RA Rosenthal A., Thierauch K.H.;
RT "Differential gene expression by endothelial cells in distinct
RT angiogenic states.";
RL Eur. J. Biochem. 267:2820-2830(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

```
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@sib-sib.ch).
CC CC -----
CC DR EMBL: AF149118; AAD34012.1; -
CC DR EMBL: AF304446; AAG29823.1; -
CC DR MEROPS: M12.222; -
CC DR InterPro: IPR001762; Disintegrin.
CC DR InterPro: IPR002870; Pep_M12B_propep.
CC DR InterPro: IPR001590; Reprolysin.
CC DR InterPro: IPR000884; TSP1.
CC DR InterPro: IPR000130; Zn_Mtpeptdse.
CC DR Pfam: PF01421; Reprolysin; 2.
CC DR Pfam: PF01562; Pep_M12B_propep; 2.
CC DR SMART: SM00209; TSP1; 3.
CC DR PROSITE: PS00215; ADAM_MEPRO; 1.
CC DR PROSITE: PS00142; ZINC_PROTEASE; 1.
CC DR PROSITE: PS00092; TSP1; 2.
CC DR PROSITE: PS00427; DISINTEGRIN_1; FALSE NEG.
CC KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
CC Repeat; Extracellular matrix; Heparin-binding.
CC FT SIGNAL 1 54 POTENTIAL.
CC FT PROPEP 55 252 BY SIMILARITY.
CC FT CHAIN 253 967 ADAMTS-1.
CC FT SITE 205 205 CYSTEINE SWITCH (POTENTIAL).
CC FT METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT ACT_SITE 402 402 BY SIMILARITY.
CC FT METAL 405 405 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT METAL 411 411 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT DOMAIN 476 558 DISINTEGRIN-LIKE.
CC FT DOMAIN 559 615 TSP TYPE-1 1.
CC FT DOMAIN 616 724 CYS-RICH.
CC FT DOMAIN 725 857 SPACER.
CC FT DOMAIN 858 907 TSP TYPE-1 2.
CC FT DOMAIN 908 967 TSP TYPE-1 3.
CC FT DOMAIN 194 198 POLY-ARG.
CC FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 720 720 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 764 764 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 945 945 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CONFLICT 21 21 I -> V (IN REF. 2).
CC FT CONFLICT 26 31 KFRSQ -> RSRGSL (IN REF. 2).
CC FT CONFLICT 49 49 V -> A (IN REF. 2).
CC FT CONFLICT 72 72 R -> P (IN REF. 2).
CC FT CONFLICT 79 79 L -> TR (IN REF. 2).
CC FT CONFLICT 249 249 TMLV -> NLK (IN REF. 2).
CC FT CONFLICT 262 265 S -> F (IN REF. 2).
CC FT CONFLICT 607 607 L -> V (IN REF. 2).
CC FT CONFLICT 936 936 I -> T (IN REF. 2).
CC FT CONFLICT 962 962 I -> T (IN REF. 2).
CC SQ SEQUENCE 967 AA; 105705 MW; P93C864F6DCDB4CF CRC64;
CC Query Match 1.8%; Score 17; DB 1; Length 967;
CC Best Local Similarity 100.0%; Pred. No. 8.2e-09;
CC Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC QY 356 AFTTAHELGHVFNMPHD 372
CC Db 396 AFTTAHELGHVFNMPHD 412
CC RESULT 3
CC ATSL_MOUSE STANDARD; PRT; 968 AA.
CC AC P97857; O54768;
CC DT 30-MAY-2000 (Rel. 39, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
```

```
DT DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
GN with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
DE ADAMTS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
OX RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RX MEDLINE=98110583; PubMed=9441751;
RA Kuno K., Lizaola H., Ohno S., Matsushima K.;
RT "The exon/intron organization and chromosomal mapping of the mouse
RT ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";
RL Genomics 46:466-471(1997).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=97150761; PubMed=8995297;
RA Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.,
RA Matsushima K.;
RT "Molecular cloning of a gene encoding a new type of metalloproteinase-
RT disintegrin family protein with thrombospondin motifs as an
RT inflammation associated gene.";
RL J. Biol. Chem. 272:556-562(1997).
RN [3]
RN CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.
RX MEDLINE=99303657; PubMed=10373500;
RA Kuno K., Terashima Y., Matsushima K.;
RT "ADAMTS-1 is an active metalloproteinase associated with the
RT extracellular matrix.";
RL J. Biol. Chem. 274:18821-18826(1999).
RN [4]
RN FUNCTION.
RX MEDLINE=20389568; PubMed=10930576;
RA Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.,
RA Ohno H., Matsushima K.;
RT "ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";
RL FEBS Lett. 478:241-245(2000).
RN [5]
RN FUNCTION, AND INDUCTION.
RX MEDLINE=20243757; PubMed=10781075;
RA Rodker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W.,
RA Richards J.S.;
RT "Progestosterone-regulated genes in the ovulation process: ADAMTS-1 and
RT cathepsin L proteases.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).
CC -!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: CLEAVES AGGECAN AT THE 1691-GLU-|-LEU-1692
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -!- COPACITOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX.
CC -!- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY
CC INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY
CC LIPOPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA
CC CELLS OF PREOVULATORY FOLLICLES.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 7.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AB001735; BAA24501.1; ALT.INT.
DR EMBL; D67076; BAA11088.1; ALT.FRAME.
DR MEROPS; M12.222; -.
DR MGD; MGI:109249; Adams1.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep.M12B.propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn.Mtpeptdse.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep.M12B.propep; 1.
DR SMART; SM00209; TSPl; 3.
DR PROSITE; PS02015; ADAM.MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS00092; TSPl; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 48
FT PROPEP 49 253
FT CHAIN 254 968
FT SITE 206 206 ADAMTS-1.
FT METAL 402 402 CYSTEINE SWITCH (POTENTIAL).
FT ACT_SITE 403 403 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 406 406 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 477 559 DISINTEGRIN-LIKE.
FT DOMAIN 560 617 TSP TYPE-1.
FT DOMAIN 618 725 CYS-RICH.
FT DOMAIN 726 850 SPACER.
FT DOMAIN 851 909 TSP TYPE-1 2.
FT DOMAIN 910 968 TSP TYPE-1 3.
FT DOMAIN 195 199 POLY-ARG.
FT CARBOHYD 548 548 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 721 721 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 765 765 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 783 783 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 946 946 N-LINKED (GLCNAC. .) (POTENTIAL).
FT MUTAGEN 403 406 E->O: LOSS OF ACTIVITY.
FT CONFLICT 335 335 N -> S (IN REF. 2).
FT CONFLICT 425 425 T -> S (IN REF. 2).
SQ SEQUENCE 968 AA; 105841 MW; 42EBDA55499FB6C1 CRC64;

Query Match 1.8%; Score 17; DB 1; Length 968;
Best Local Similarity 100.0%; Pred. No. 8.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 AFTTAHELGHVFNMPHD 372
Db 397 AFTTAHELGHVFNMPHD 413
|||||
RESULT 4
AT54_RAT STANDARD; PRT; 630 AA.
AC Q9ESP7; Q9ESP8; Q9ESP6;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 4) (ADAM-TS 4) (Aggrecanase 1)
DE (Fragment).
DE GN ADAMTS4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar; TISSUE=Brain;
RX MEDLINE=20415831; PubMed=10961658;
RA Satoh K., Suzuki N., Yokota H.;
RT "ADAMTS-4 (a disintegrin and metalloproteinase with thrombospondin
RT motifs) is transcriptionally induced in beta-amyloid treated rat
RT astrocytes.";
RL Neurosci. Lett. 289:177-180(2000).
CC -!- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES.
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393
CC site.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: ONLY EXPRESSED IN BRAIN.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AB042272; BAB16474.1; -.
DR EMBL; AB042271; BAB16473.1; -.
DR EMBL; AB042273; BAB16475.1; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn.Mtpeptdse.
DR Pfam; PF00090; tsp_1; 2.
DR Pfam; PF01421; Reprolysin; 1.
DR SMART; SM00209; TSPl; 1.
DR PROSITE; PS00215; ADAM.MEPRO; 1.
DR PROSITE; PS00092; TSPl; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
KW Extracellular matrix.
FT NON_TER 1 1 BY SIMILARITY.
FT PROPEP <1 5 ADAMTS-4.
FT CHAIN 6 630 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 154 154 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 155 155 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 158 158 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 164 164 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 233 303 DISINTEGRIN-LIKE.
FT DOMAIN 316 367 TSP TYPE-1.
FT DOMAIN 368 478 CYS-RICH.
FT DOMAIN 479 630 SPACER.
FT DOMAIN 40 45 POLY-ALA.
FT CARBOHYD 96 96 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 474 474 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 630 AA; 68384 MW; 63A428753167C7EF CRC64;

Query Match 1.4%; Score 13; DB 1; Length 630;
Best Local Similarity 100.0%; Pred. No. 7.6e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 311 DTAIFTRQDLGG 323
Db 104 DTAIFTRQDLGG 116
|||||
```

```

RESULT 5
ID ATSA_HUMAN STANDARD; PRT; 837 AA.
AC 075173; Q9UN83;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
DE (ADMP-1).
GN ADAMTS4 OR KIAA0688.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99286303; PubMed=10356395;
RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
RA Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
RA Rockwell A., Tang F., Duke J.L., Solomon K., George H., Bruckner R.,
RA Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
RA Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
RA Trzaskos J.M., Arner E.C.;
RT "Purification and cloning of aggrecanase-1: a member of the ADAMTS
RT family of proteins.";
RL Science 284:1664-1666(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Sawaji Y., Nagase H., Saklatvala J., Clark A.R.;
RT "ADAMTS-4 genomic locus.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=20400518; PubMed=10827174;
RA Tortorella M.D., Pratta M.A., Liu R.-Q., Abbaszade I., Ross H.,
RA Burn T.C., Arner E.C.;
RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
RT aggrecan substrate recognition and cleavage.";
RL J. Biol. Chem. 275:25791-25797(2000).
CC -1- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE
CC CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN
CC ALZHEIMER'S DISEASE.
CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393
CC site.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED
CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
CC -1- INDUCTION: BY INTERLEUKIN-1.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO AS ADAMTS2.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AB014588; BAA31663.1; -
CC EMBL; AF148213; AAD41494.1; -
CC EMBL; AY044847; AAL02262.1; -
CC MEROPS; M12.221; -
CC GENE; HGNC:220; ADAMTS4.
CC MIM; 603876; -
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; Zn_Mtpeptidse.
CC Pfam; PF00090; tsp_1; 1.
CC SMART; SM00209; TSP1; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS00092; TSP1; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Extracellular matrix.
FT SIGNAL 1 51 POTENTIAL.
FT PROPEP 52 212
FT CHAIN 213 837
FT SITE 194 194
FT METAL 361 361
FT ACT_SITE 361 362
FT METAL 365 365
FT METAL 371 371
FT DOMAIN 437 519
FT DOMAIN 520 576
FT DOMAIN 577 685
FT DOMAIN 686 837
FT DOMAIN 247 252
FT CARBOHYD 68
FT CONFLICT 77 77
FT CONFLICT 626 626
FT CONFLICT 682 682
SQ SEQUENCE 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;
Query Match 1.4%; Score 13; DB 1; Length 837;
Best Local Similarity 100.0%; Pred. No. 9.7e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 311 DTALFTRODLG 323
DB 311 DTALFTRODLG 323
-----
RESULT 6
ID ATSA_HUMAN STANDARD; PRT; 1629 AA.
AC Q9P2N4; Q9NR29;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).
GN ADAMTS9 OR KIAA1312.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE=Fetal;
RX MEDLINE=20396138; PubMed=10936055;

```


modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AF175282; AAF25805.1; -
 HSP: P34179; LIAG.
 MEROPS: M12.226; -
 MGI: 1353468; Adamts8.
 InterPro: IPR001762; Disintegrin.
 InterPro: IPR001590; Reprolysin.
 InterPro: IPR000884; TSPI.
 InterPro: IPR000130; zn_Mtpeptdse.
 Pfam: PF00090; tsp_1; 2.
 Pfam: PF01421; Reprolysin; 1.
 SMART: SM00209; TSPI; 2.
 PROSITE: PS00142; ZINC_PROTEASE; 1.
 PROSITE: PS00215; ADAM_MEROPS; 1.
 PROSITE: PS0092; TSPI; 1.
 PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
 Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Extracellular matrix; Heparin-binding.
 FT SIGNAL 1 28 POTENTIAL.
 FT PROPEP 29 228 BY SIMILARITY.
 FT CHAIN 229 905 ADAMTS-8.
 FT METAL 378 378 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 379 379 BY SIMILARITY.
 FT METAL 382 382 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 453 541 DISINTEGRIN-LIKE.
 FT DOMAIN 542 598 TSP TYPE-1 1.
 FT DOMAIN 599 705 CYS-RICH.
 FT DOMAIN 706 847 SPACER.
 FT DOMAIN 848 905 TSP TYPE-1 2.
 FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 506 506 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 905 AA; 98879 MW; 124D4132B33A0CAE CRC64;
 Query Match 1.3%; Score 12; DB 1; Length 905;
 Best Local Similarity 100.0%; Pred. No. 0.0011;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 603 WPKYGVSPRD 614
 |||||
 DB 626 WPKYGVSPRD 637
 . RESULT 8
 AT55_HUMAN STANDARD; PRT; 930 AA.
 AC Q9UNA0; Q9UKP2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
 DE (ADMP-2) (ADAM-TS 11).
 GN ADAMTS5 OR ADMP2 OR ADAMTS11.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=99367476; PubMed=10438522;
 RA Abbszade I., Liu R.-Q., Yang F., Rosenfeld S.A., Ross O.H.,
 RA Link J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M.,
 RA Wynn R., Duke J.L., George H.J., Hillman M.C. Jr., Murphy K.,
 RA Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner R., Nagase H.,
 RA Ito Y., Newton R.C., Magolda R.L., Trzaskos J.M., Hollis G.F.,
 RA Arner E.C., Burn T.C.;

"Cloning and characterization of ADAMTS11, an aggrecanase from the
 ADAMTS family.";
 J. Biol. Chem. 274:23443-23450(1999).
 [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=20289799; PubMed=10830953;
 RX Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Tokoki Y., Choi D.-K., Soeda E.,
 RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
 RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Blocker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Reinhardt R., Yaspo M.-L.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.,
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 [3]
 RN SEQUENCE OF 413-930 FROM N.A.
 RP TISSUE=Fetal brain;
 RC MEDLINE=9935124; PubMed=10464288;
 RX Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
 RA "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
 RT zinc metalloproteases.";
 RL J. Biol. Chem. 274:25555-25563(1999).
 CC -1- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
 INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
 DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
 PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
 CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393
 site.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 matrix (by similarity).
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA
 BUT ALSO IN OTHER TISSUES, SUCH AS HEART AND BRAIN, AND ALSO
 CC CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE, FROM AN
 CC CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN
 CC ARTHRITIC PATIENT.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: AF142099; AAD49577.1; -
 EMBL: AP001698; BAA95504.1; -
 EMBL: AP001697; BAA95503.1; -
 EMBL: AF141293; AAF02493.1; -
 HSP: Q9PW35; 1BUD.
 MEROPS: M12.225; -
 Genew: HGNC:221; ADAMTS5.
 MIM: 605007; -
 InterPro: IPR001762; Disintegrin.
 InterPro: IPR002870; Pep_M12B_propep.
 InterPro: IPR001590; Reprolysin.
 InterPro: IPR000884; TSPI.
 InterPro: IPR000130; zn_Mtpeptdse.
 Pfam: PF00090; tsp_1; 2.
 Pfam: PF01421; Reprolysin; 1.

DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2)
DE (METH-8)
GN ADAMTS8 OR METH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=99367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
RA Lombardo M., Iruela-Arispe M.L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
RT family of proteins with angio-inhibitory activity.";
RL J. Biol. Chem. 274:23349-23357(1999).
RN [2]
RP SEQUENCE OF 195-440 FROM N.A.
RX MEDLINE=20079168; PubMed=10610729;
RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS8, a novel metalloproteinase of the ADAM-TS family located on
RT mouse chromosome 9 and human chromosome 11.";
RL Genomics 62:312-315(1999).
CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER
CC EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND
CC KIDNEY.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF060153; AAD48081.1; -.
DR EMBL; AF175283; AAF25806.1; -.
DR HSSP; P34179; 1IA6.
DR MEROPS; M12.226; -.
DR Genew; HGNC:224; ADAMTS8.
DR MIM; 605175; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTPeptdse.
DR Pfam; PF00090; tsp_1; 2.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPl; 2.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00092; TSPl; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloproteinase; Zinc; Signal; Glycoprotein; zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 214 BY SIMILARITY.
FT CHAIN 215 890 ADAMTS-8.
FT METAL 364 364 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 365 365 BY SIMILARITY.

FT METAL 368 368 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 374 374 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 439 526 DISINTEGRIN-LIKE.
FT DOMAIN 527 583 TSP TYPE-1 1.
FT DOMAIN 584 690 CYS-RICH.
FT DOMAIN 691 832 SPACER.
FT DOMAIN 833 890 TSP TYPE-1 2.
FT DOMAIN 202 205 POLY-PRO.
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 600 600 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 195 195 E -> R (IN REF. 2).
FT CONFLICT 413 440 YLFEILDGGHGDCLLDAPGAALPLPTGL -> FSGCHLQGW
FT IHEFKYLCVKVSELKCDLMP (IN REF. 2).
SQ SEQUENCE 890 AA; 96671 MW; 57D70EE03D5739D3 CRC64;
Query Match 1.1%; Score 10; DB 1; Length 890;
Best Local Similarity 100.0%; Pred. No. 0.13; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;
QY 528 CSRTCGGGVQ 537
| | | | | | | | | |
Db 539 CSRTCGGGVQ 548
RESULT 11
AT10_HUMAN
ID AT10_HUMAN STANDARD; PRT; 1077 AA.
AC Q9H324;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).
GN ADAMTS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Apte S.S.;
RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
RT thrombospondin type 1 repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF163762; AAG35563.1; -.
DR MEROPS; M12.235; -.
DR Genew; HGNC:13201; ADAMTS10.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTPeptdse.

```

DR Pfam; PF00090; tsp_1; 5.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPl; 5.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSPl; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT NON_TER 1 1
FT PROPEP 1 207 BY SIMILARITY.
FT CHAIN 208 1077 ADAMTS-10.
FT METAL 366 366 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 367 367 BY SIMILARITY.
FT METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 376 376 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 434 520 DISINTEGRIN-LIKE.
FT DOMAIN 578 679 CYS-RICH.
FT DOMAIN 680 802 SPACER.
FT DOMAIN 521 577 TSP TYPE-1 1.
FT DOMAIN 799 860 TSP TYPE-1 2.
FT DOMAIN 862 918 TSP TYPE-1 3.
FT DOMAIN 922 976 TSP TYPE-1 4.
FT DOMAIN 981 1031 TSP TYPE-1 5.
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 714 714 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 769 769 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 866 866 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1077 AA; 118072 MW; 3914DE1DCBFB587 CRC64;

Query Match 1.1%; Score 10; DB 1; Length 1077;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 RYVETLVVAD 227
| | | | | | | | | |
Db 213 RYVETLVVAD 222

RESULT 12
AT52_BOVIN STANDARD; PRT; 1205 AA.
AC P79331;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI).
GN ADAMTS2 OR NPI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin.
RX MEDLINE-9725960; PubMed=9122202;
RA Colige A., Li S.W., Sieron A.L., Nusgens B.V., Prockop D.J.,
RA Lapiere C.M.;
RT "cDNA cloning and expression of bovine procollagen I N-proteinase; a
RT new member of the superfamily of zinc-metalloproteinases with binding
RT sites for cells and other matrix components."
RL Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE-95348096; PubMed=7622483;
RA Colige A., Beschin A., Samyn B., Goebels Y., Van Beeumen J.,

```

```

RA Nusgens B.V., Lapiere C.M.;
RT "Characterization and partial amino acid sequencing of a 107-kDa
RT procollagen I N-proteinase purified by affinity chromatography on
RT immobilized type XIV collagen."
RL J. Biol. Chem. 270:16724-16730(1995).
CC -!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
CC COLLAGEN BIOSYNTHESIS.
CC -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
CC alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains
CC at Ala-1-Gln.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
CC COLLAGEN TYPE XIV.
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVEL
CC IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS
CC AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS
CC WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF DERMATOSPARAXIS, A
CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN
CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN
CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X96389; CAA65253.1; -
DR MEROPS; M12.301; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Repolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF00090; tsp_1; 4.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPl; 4.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSPl; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 28 POTENTIAL.
FT PROPEP 29 253 BY SIMILARITY.
FT CHAIN 254 1205 ADAMTS-2.
FT METAL 402 402 ZINC (CATALYTIC) (POTENTIAL).
FT ACT_SITE 403 403 BY SIMILARITY.
FT METAL 406 406 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 474 554 DISINTEGRIN-LIKE.
FT DOMAIN 555 611 TSP TYPE-1 1.
FT DOMAIN 612 716 CYS-RICH.
FT DOMAIN 717 845 SPACER.
FT DOMAIN 846 905 TSP TYPE-1 2.
FT DOMAIN 906 968 TSP TYPE-1 3.

```

```

FT DOMAIN 969 1024 TSP TYPE-1 4.
FT SITE 685 CELL ATTACHMENT SITE (POTENTIAL).
FT DOMAIN 31 POLY-ALA.
FT DOMAIN 177 180 POLY-GLU.
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 943 943 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 987 987 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1025 1025 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1092 1092 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1144 1144 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1205 AA; 133887 MW; 7B5B232A45320371 CRC64;

Query Match 1.1%; Score 10; DB 1; Length 1205;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 DKCGVCGGDN 679
Db 704 DKCGVCGGDN 713

RESULT 13
AT52_HUMAN
ID AT52_HUMAN STANDARD; PRT; 1205 AA.
AC O15072; Q9BXZ8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ADAMTS-3 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 3) (ADAM-TS 3) (Procollagen II
DE amino-propeptide processing enzyme) (Procollagen II N-proteinase) (PC
DE II-NP)
GN ADAMTS3 OR KTA0366.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-227 FROM N.A.
RX MEDLINE=21402912; PubMed=11408482;
RA Fernandes R.J., Hirohata S., Engle J.M., Colige A., Cohn D.H.,
RA Eyre D.R., Apte S.S.;
RT "Procollagen II amino propeptide processing by ADAMTS-3. Insights on
RT dermatosparaxis."
RL J. Biol. Chem. 276:31502-31509(2001).
RN [2]
RP SEQUENCE OF 5-1205 FROM N.A.
RX TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 4:141-150(1997).
CC -!- FUNCTION: Cleaves the propeptides of type II collagen prior to
CC fibril assembly. Does not act on types I and III collagens.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: Found in cartilage and skin.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.

```

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF247668; AAK28400.1; -
DR EMBL; AB002364; BAA20821.1; -
DR MEROPS; M12.220; -
DR Genew; HGNC:219; ADAMTS3.
DR MTM; 605011; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_MTPetdse.
DR Pfam; PF00090; tsp.1; 4.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPI; 4.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS0092; TSPI; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloproteinase; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 249 BY SIMILARITY.
FT CHAIN 250 1205 ADAMTS-3.
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 399 399 BY SIMILARITY.
FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 408 408 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 470 550 DISINTEGRIN-LIKE.
FT DOMAIN 551 607 TSP TYPE-1 1.
FT DOMAIN 608 712 CYS-RICH.
FT DOMAIN 713 844 SPACER.
FT DOMAIN 845 902 TSP TYPE-1 2.
FT DOMAIN 903 965 TSP TYPE-1 3.
FT DOMAIN 966 1017 TSP TYPE-1 4.
FT DOMAIN 246 249 POLY-ARG.
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 475 475 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1205 AA; 135570 MW; EB07B286FC85FB87 CRC64;

Query Match 1.1%; Score 10; DB 1; Length 1205;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 670 DKCGVCGGDN 679
Db 700 DKCGVCGGDN 709

RESULT 14
AT52_HUMAN
ID AT52_HUMAN STANDARD; PRT; 1211 AA.
AC Q95450;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI)
DE (Procollagen I/II amino-propeptide processing enzyme).

```

GN ADAMTS2 OR PCINP OR PCPNI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI), AND DISEASE.
RC TISSUE=Skin;
RA MEDLINE=99347935; PubMed=10417273;
RA Colige A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,
RA Wertelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,
RA Byers P.H., Lapiere C.M., Prockop D.J., Nussgens B.V.;
RA "Human Ehlers-Danlos syndrome type VII c and bovine dermatosparaxis
RA are caused by mutations in the procollagen I N-proteinase gene.";
RL Am. J. Hum. Genet. 65:308-317(1999).
CC -1- FUNCTION: Cleaves the propeptides of type I and II collagen prior
CC to fibril assembly. Does not act on type III collagen. May also
CC play a role in development that is independent of its role in
CC collagen biosynthesis.
CC -1- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
CC alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains
CC at Ala-1-Gln.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
CC COLLAGEN TYPE XIV (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI;
CC ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-
CC PROCOLLAGEN PEPTIDASE ACTIVITY.
CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON
CC AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos
CC syndrome type VIIC (EDS-VIIC), a recessively inherited connective-
CC tissue disorder characterized clinically by severe skin fragility
CC and joint hypermobility and biochemically by the presence in skin
CC of procollagen incompletely processed at the N-terminus.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ003125; CAA05880.1; -
DR MEROPS; M12.301; -
DR Genew; HGNC:218; ADAMTS2.
DR MIN; 604539; -
DR MIN; 225410; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF00090; tsp-1; 4.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 4.
DR PROSITE; PS02015; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSP1; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolyase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;

KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;
KW Alternative splicing; Ehlers-Danlos syndrome.
FT SIGNAL 1 29
FT PROPEP 30 253
FT CHAIN 1211
FT CHAIN 254 1211
FT METAL 408 408
FT ACT_SITE 409 409
FT METAL 412 412
FT METAL 418 418
FT DOMAIN 480 560
FT DOMAIN 561 617
FT DOMAIN 618 722
FT SITE 691 693
FT DOMAIN 723 851
FT DOMAIN 852 911
FT DOMAIN 912 974
FT DOMAIN 975 1030
FT DOMAIN 40 43
FT DOMAIN 185 188
FT CARBOHYD 112 112
FT CARBOHYD 251 251
FT CARBOHYD 949 949
FT CARBOHYD 993 993
FT CARBOHYD 1031 1031
FT CARBOHYD 1098 1098
FT CARBOHYD 1145 1145
FT CARBOHYD 1150 1150
FT VARSPLIC 544 566
FT VARSPLIC 567 1211
FT SEQUENCE 1211 AA; 134722 MW; BECEF25C33CAD2D CRC64;
SQ
Query Match 1.1%; Score 10; DB 1; Length 1211;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 670 DKCGVCGGDN 679
Db 710 DKCGVCGGDN 719
RESULT 15
AT56.HUMAN
ID AT56.HUMAN STANDARD; PRT; 860 AA.
AC Q9URP5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-6 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 6) (ADAM-TS 6) (ADAM-TS6).
GN ADAMTS6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=99395124; PubMed=10464288;
RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RA "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RA Zinc Metalloproteases";
RL J. Biol. Chem. 274:25555-25563(1999).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN PLACENTA AND BARELY
CC DETECTABLE IN A NUMBER OF OTHER TISSUES.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

```

CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF140674; RAD56357.1; -.
DR HSSP; P34179; ILAG.
DR MEROPS; M12.230; -.
DR Genew; HGNC:222; ADAMTS6.
DR MIM; 605008; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Repolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR00130; zn_M1peptdse.
DR Pfam; PF00090; tsp.1; 1.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPI; 1.
DR PROSITE; PS0215; ADAM_MPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00092; TSPI; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 21
FT PROPEP 22 244
FT CHAIN 245 860
FT METAL 403 403
FT ACT_SITE 404 404
FT METAL 407 407
FT METAL 413 413
FT DOMAIN 453 509
FT DOMAIN 510 566
FT DOMAIN 567 668
FT DOMAIN 669 795
FT DOMAIN 796 852
FT DOMAIN 68 71
FT DOMAIN 662 665
FT CARBOHYD 99 99
FT CARBOHYD 172 172
FT CARBOHYD 222 222
FT CARBOHYD 234 234
FT CARBOHYD 676 676
FT CARBOHYD 843 843
SQ SEQUENCE 860 AA; 97098 MW; E57213015DECB2C5 CRC64;

Query Match 0.9%; Score 9; DB 1; Length 860;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 528 CSRTCGGV 536
Db- 522 CSRTCGGV 530
|||||||

```

Search completed: April 29, 2003, 17:23:57
Job time : 19 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 22.6233 Seconds
(without alignments)
3378.970 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_583

Perfect score: 2042
Sequence: 1 FVSIPRYVETLVVADESMVK.....SCNLEPCPSSASGKSFREQ 371

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

1: sp-archaea.*

2: sp-bacteria.*

3: sp-fungi.*

4: sp-human.*

5: sp-invertebrate.*

6: sp-mammal.*

7: sp-mhc.*

8: sp-organelle.*

9: sp-phage.*

10: sp-plant.*

11: sp-rodent.*

12: sp-virus.*

13: sp-vertebrate.*

14: sp-unclassified.*

15: sp-rvirus.*

16: sp-bacteriap.*

17: sp-archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2042	100.0	950	Q8TE58	Q8TE58 homo sapien
2	923.5	45.2	2165	Q19791	Q19791 caenorhabdi
3	796	39.0	269	Q9GL54	Q9GL54 oryctolagus
4	796	39.0	1054	Q9W493	Q9W493 drosophila
5	703	34.4	1688	Q8SX80	Q8SX80 drosophila
6	670.5	32.8	1229	Q9VF61	Q9VF61 drosophila
7	654	32.0	1072	Q8TE57	Q8TE57 homo sapien
8	643.5	31.5	1095	Q8TE56	Q8TE56 homo sapien
9	638.5	31.3	1081	Q8TE60	Q8TE60 homo sapien
10	617.5	30.2	1223	Q8WX58	Q8WX58 homo sapien
11	617.5	30.2	1223	Q8TE55	Q8TE55 homo sapien
12	616	30.2	1159	Q8TEY8	Q8TEY8 homo sapien
13	606	29.7	1207	Q8TE59	Q8TE59 homo sapien
14	558	27.3	192	Q95N24	Q95N24 equus caball
15	540	26.4	1427	Q96L37	Q96L37 homo cabal
16	518	25.4	187	Q95N23	Q95N23 equus cabal

17	396.5	19.4	1062	5	Q19204	Q19204 caenorhabdi
18	393	19.2	872	5	Q22580	Q22580 caenorhabdi
19	377.5	18.5	1444	5	Q17591	Q17591 caenorhabdi
20	361.5	17.7	1091	5	Q9W126	Q9W126 drosophila
21	357.5	17.5	790	5	Q8T458	Q8T458 drosophila
22	303	14.8	108	6	Q9GLK6	Q9GLK6 oryctolagus
23	298	14.6	117	11	Q8VHK4	Q8VHK4 mus musculu
24	237.5	11.6	364	4	Q9UGQ1	Q9UGQ1 homo sapien
25	231.5	11.3	509	5	Q9NDL4	Q9NDL4 caenorhabdi
26	231.5	11.3	509	5	Q20930	Q20930 caenorhabdi
27	231.5	11.3	899	13	Q8UVF1	Q8UVF1 coturnix co
28	226	11.1	610	13	Q93523	Q93523 bothrops ja
29	226	11.1	610	13	Q8QG88	Q8QG88 bothrops ja
30	224.5	11.0	477	13	Q98SP2	Q98SP2 bothrops ja
31	224	11.0	922	13	Q8UVF2	Q8UVF2 coturnix co
32	223.5	10.9	609	13	Q9W6M5	Q9W6M5 agkistrodon
33	221	10.8	482	13	Q9PVK9	Q9PVK9 agkistrodon
34	220.5	10.8	465	13	Q9IAX7	Q9IAX7 agkistrodon
35	220.5	10.8	479	13	Q9PWJ0	Q9PWJ0 agkistrodon
36	218.5	10.7	419	13	Q92043	Q92043 crotalus at
37	216.5	10.6	476	13	Q9YI19	Q9YI19 agkistrodon
38	214.5	10.5	479	13	Q9PW78	Q9PW78 agkistrodon
39	213	10.4	483	13	Q9IAB0	Q9IAB0 agkistrodon
40	213	10.4	549	13	Q90500	Q90500 echis pyram
41	212	10.4	4123	4	Q75851	Q75851 homo sapien
42	210.5	10.3	612	13	Q8UVG0	Q8UVG0 bothrops er
43	210	10.3	610	13	Q9YI20	Q9YI20 agkistrodon
44	209.5	10.3	617	13	Q90499	Q90499 echis pyram
45	208.5	10.2	609	13	Q90282	Q90282 crotalus at

ALIGNMENTS

RESULT 1

Q8TE58	PRELIMINARY;	PRT;	950 AA.
AC	Q8TE58;		
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Metalloprotease disintegrin 15 with thrombospondin domains.		
GN	ADAMTS15.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21856482; PubMed=11867212;		
RA	Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,		
RA	Lopez-Otin C.;		
RT	seven novel human ADAMTSs, a family of metalloproteinases with		
RT	disintegrin and thrombospondin-1 domains.;		
RL	Gene 283:49-62(2002).		
DR	EMBL; AJ315733; CAC86014.1; -		
DR	Integrin; Protease.		
SQ	SEQUENCE 950 AA; 103286 MW; 5DFBE18285CCCC3B CRC64;		

Query Match 100.0%; Score 2042; DB 4; Length 950;
Best Local Similarity 100.0%; Pred. No. 1.1e-195;
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	FVSIPRYVETLVVADESMVKFHGADLEHYLLTLAARLYRHPSILNFINVVKVLL 60
Db	213	FVSIPRYVETLVVADESMVKFHGADLEHYLLTLAARLYRHPSILNFINVVKVLL 272
QY	61	RDRDGPVKVNGAATLNFCAWOKKLNKVSQKHPYWDYTAITLFTRODLGATTCDFLGM 120
Db	273	RDRDGPVKVNGAATLNFCAWOKKLNKVSQKHPYWDYTAITLFTRODLGATTCDFLGM 332
QY	121	ADVGTMCDFKRSVCIEDDGLPFAFTTAHELGHVFNMPHDNVKVCYEEVFGKLRANHMSP 180

```

Db 333 ADYGTWCDPKRSVTEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEYFGKLRANHMSP 392
QY 181 TLIOIDRANPWSACSAAITDLDGSHGDCILDDQSPKSPISLPEDLPAGASYTLISOCELA 240
Db 393 TLIOIDRANPWSACSAAITDLDGSHGDCILDDQSPKSPISLPEDLPAGASYTLISOCELA 452
QY 241 GVSGKPCPYMYCTKLMCTGKAGQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 300
Db 453 GVSGKPCPYMYCTKLMCTGKAGQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 512
QY 301 HRYDGSWAKWDYGPSCSRTCCGGVQLARQCTNPTPANGKYCEGVRYKYSNLEPCPS 360
Db 513 HRYDGSWAKWDYGPSCSRTCCGGVQLARQCTNPTPANGKYCEGVRYKYSNLEPCPS 572
QY 361 SASGKSFRREQ 371
Db 573 SASGKSFRREQ 583

RESULT 2
Q19791 PRELIMINARY; PRT; 2165 AA.
AC Q19791; Q27524;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE F25H8.3 protein.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OC Rhabditiida; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP Gajasty S.;
RA Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
DR EMBL; Z69361; CAA93288.1;
DR EMBL; Z69360; CAA93288.1; JOINED.
DR EMBL; Z69360; CAA93287.1;
DR EMBL; Z69361; CAA93287.1; JOINED.
DR HSSP; P15167; 1DTH.
DR MEROPS; M12.135; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTPeptdse.
DR Pfam; PF01421; Reprolysin; 1.
DR SMART; SM00209; TSP1; 18.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSP1; 6.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 2165 AA; 244397 MW; FCC3DA8AA9C4888 CRC64;

Query Match 45.2%; Score 923.5; DB 5; Length 2165;
Best Local Similarity 45.4%; Pred. No. 4.3e-83;
Matches 177; Conservative 60; Mismatches 126; Indels 27; Gaps 7;

```

```

QY 7 YVELTVVADESVMKPHGADLEHYLLTLLAARLYRHPHSILNPINIVVVKVLLLRDRDSG 66
Db 281 YVELTVVADEKMEYHGRSLDEYVLTSTVASIYRHSQSLRASINVVVVKLVLTENAG 340
QY 67 PKVTGNAATLNFCAWQKKLNKVDKPEYWDTAITFTRODLGAT-TCDTLGMADVGT 125
Db 341 PRITQNAOCTLODFCRWOQYNDPDDSSVQHHDAILLTRKIDICRSQCKDTLGLAELGT 400
QY 126 MCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEV-----FGKL 172
Db 401 MCDMQKSCAIIEDNGLSAAFTIAHELGHVFSIPHDERKCKSTYMPVNVKVCFOSTKEDKT 460
QY 173 RAN---HMSPTLIQIDRANPWSACSAAITDLDGSHG---DCLLDQSPKSPISPE---- 223
Db 461 OFQNNPHIMAPTLEYNTHPWSWSPCSAGMLERFLENNGOTQCLFDQPVERRYEDYFVR 520
QY 224 DLPGASYTLISOCELAFAFGVSGKPCPYMYCTKLMCTGKAGQVQVQVQVQVQVQVQV 283
Db 521 DEPGKKYDAHQCKKFVFGPASELCPYMPTCRLWCATFYGSMGCRTOHMPWADGTPCDE 580
QY 284 GK--ICLKAGACVERHNLNKHVRDGSWAKWDYGPSCSRTCCGGVQLARQCTNPTPANGK 341
Db 581 SRSMFCHHGACVRLAPESLTKIDGQWDMRWGECSTCCGGVQKGLRDCDSPKPRNGK 640
QY 342 YCEGVRYKYSNLEPCPSASGKSFRREQ 371
Db 641 YCVGQREYRSCNTOECPWDT--QPYREVQ 668

RESULT 3
Q9GL54 PRELIMINARY; PRT; 269 AA.
ID Q9GL54;
AC Q9GL54;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Aggrecanase-2 (Fragment).
GN ADAMTS-11.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Coad D.L., Coad M.E.;
RT "Molecular cloning of lapine aggrease-2 (ADAMTS-11) from articular
RT chondrocytes."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF317415; AAG33062.1; -.
DR HSSP; Q9PW35; 1BUD.
DR MEROPS; M12.225; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTPeptdse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 1.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSP1; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
FT NON_TER 1
FT NON_TER 269
SQ SEQUENCE 269 AA; 29193 MW; 97A1CA80B33452FA CRC64;

Query Match 39.0%; Score 796; DB 6; Length 269;
Best Local Similarity 53.8%; Pred. No. 1.7e-71;
Matches 140; Conservative 37; Mismatches 73; Indels 10; Gaps 3;

QY 117 TLGMADVGTMCDPKRSVTEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEYFGKLRANH 176
Db 1 TLGMADVGTICSPERSCAVIEDDGLHAAFTVAHEITGLLGLSHDSDKSFCEENFGSTEDKR 60

```

	DR	EURL; AF003435; AAF46065.1; -.	-.	
	DR	HSP; P15167; IATL.	IATL	
	DR	MEROPS; M12.231; -.	-.	
	DR	Flybase; FBgn0029791; CG4096.	CG4096	
	DR	InterPro; IPRO02870; Pep_M12B_propep.	Pep_M12B_propep	
	DR	InterPro; IPRO01590; Reprolysins.	Reprolysins	
	DR	InterPro; IPR000884; TSPl.	TSPl	
	DR	InterPro; IPR000130; Zn_MTpeptdse.	Zn_MTpeptdse	
	DR	pfam; PF01562; Pep_M12B_propep; 1.	Pep_M12B_propep; 1	
	DR	pfam; PF01421; Reprolysins; 1.	Reprolysins; 1	
	DR	SMART; SM00209; TSP1; 2.	TSP1; 2	
	DR	PROSITE; PS02115; ADAM_MEPRO; 1.	ADAM_MEPRO; 1	
	DR	PROSITE; PS50092; TSPI; 1.	TSPI; 1	
	DR	PROSITE; PS00142; ZINC_PROTEASE;	ZINC_PROTEASE; UNKNOWN_1.	
	SQ	SEQUENCE 1054 AA; L18616 MW; DC15455555CB6212 CRC64;		
	Query Match	39.0%; Score 796; DB 5; Length 1054;		
	Best Local Similarity	44.8%; Pred. No. le-70;		
	Matches 173; Conservative 52; Mismatches 127; Indels 34; Gaps			
QY	2	VSPRVETVLVADESMVKFEGHADLEHYLLTLATAARLYRHPISILNPINIVVKKVLLLR	G1	
DG	321	ISSPHRVETLVLDATMSAPH-RDLNGYLITIMNVSALYKDPISGNSEIIVVRRIIQD	L379	
QY	62	DRDSGGPK--VTGNAALTIRNFCAWOKKLNVSKDKPEYWDTAIFTRQDICGATTCDTIG	L119	
DG	380	EESQLQLNLTONAQRKNLDFCSWOHQHLNKNSSEKDPPHHDDVALLTRKNIC-ANNCWTLG	K348	
QY	120	MADVGTMC DPRSCSIEDGLPSAFFTAHELGHVNMPHDNVKV-CEEVFGLKRANHMM	N178	
DG	439	LANYGVGMCKPQCSCSVNEVDINGMILSHTTITHELGHNFGEFHDTAKIGHPRVGPI--VMH	A496	
QY	179	SPTL----IOIDRAPNKSACSAAITFDLSHGSDCLLODQSPXPI---SLPEDLPASGYT	J231	
DG	497	TPTFGADTLQV----CWSNCSRXYTHFLDGQGLEEC-LDDPTPLDEYNLTGELPMRYN	B551	
QY	232	LSOCELAFG-----VGSKPCCPYMOYCTKLWCAGKAQMVCOTRHFPWADGTCSEGK	C285	
DG	552	ARGQCRLFQNLTITDSEVGACSAPH-EPCSTLWC-KVNGE-CVTHMRTAPTACTLCGRNK	R606	
QY	286	LCLKGCACVERHNLNKHRYDCSWAKNDYPGPCSTRCTGGVOLARRCTNNPTPANNGKYCEG	X345	
DG	607	WCQMGKCVCRRREL--AAVGGWGMDWSWSCSSRCSGCVGSTQQRECPVPVANGGVFCIG	F664	
QY	346	VRKYRSCLNPCPSSASGSRFREEQ 371		
DG	665	ERKYYIKCRKRPCP-AEEP SFRAQQ 688		
RESULT 5				
Q8SXBO	ID	PRELIMINARY; PRt; 1688 AA.		
AC	Q8SXBO			
IC	OBSXB0:			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DG	GH16393p.			
DN	CG6107.			
OS	Drosophila melanogaster (fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxId=7227;			
FN	{1}			
RP	SEQUENCE FROM N.A.			
RC	SYRAIN-BERKELEY:			
RC	Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,			
RA	Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,			
RA	George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,			
RA	Miranda A., Mungall C.J., Nunoo J., Paciej J., Paragaa V., Park S.,			
RA	Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,			
RA	Celniker S.;			

```
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY094716; AAM11069.1; -
SQ SEQUENCE 1688 AA; 189867 MW; 48FB8DD4DE0CAAD2 CRC64;

Query Match          34.4%; Score 703; DB 5; Length 1688;
Best Local Similarity 39.3%; Pred. No. 4e-61;
Matches 151; Conservative 64; Mismatches 125; Indels 44; Gaps 15;

Qy 8 VETLVVADESVMVKGADLEHYLLTLATAARLYRHPSILNPINIVVVKVLLLRDRSGP 67
Db 457 LEVLIAVDNSMKQFHGEDLPQYILILMSIVSSIFADASIGNSIRILLVRLISL-----P 510

Qy 68 KV---TGNAAITLNFCAWQKLNKVDKHPYWDTAILTRQDLGCAT---TCDTLGMA 121
Db 511 NINDOTHSSNEMLKHFCQF---INQSYER---DTAMLTITREPICGSPVKICHMLGLA 563

Qy 122 DVGWMDPKRCSVIEDGLPSAFTTAHELGHVFNMPHDNVKVEEVEFGKLARAN---HMM 178
Db 564 ELGTVCSS-SSCSIVQDTGLPTAFTMAHELGHILNMHDDDDCKMPYVTRQNNKVLHIM 622

Qy 179 SPTLIQIDRANPWSACSAIITDPLDSDGHDCLLDQSPKPSLSP---EDLPGASYTLISQ 235
Db 623 SSVMIHMHPSWNSKSRHFVSEFLKTRDKSCL--ETSVGAHIPYGTLPGEIYSLDAQ 680

Qy 236 CELAFGVGSKPCPYMXYCTKLWC---TGKAKGMVCTRHFPWADGTSFG-EGKLCILKGA 291
Db 681 COLSEFNDGVCPTDECKRLWCNRTSGNSNEQ--CASSNLPWADGTPCGSSGHWQCRG 738

Qy 292 CVERHNLNKH----RVDGSNAKMDPYGCSRTCGGGVQLARRQCTNPTPANGKYCEGVR 347
Db 739 CVS-----NKHGYGRQVNGGMPWTFPTPCSLTCGGGVQESRRRCNPVPPENGKYCTGSR 794

Qy 348 VKYRSCNLEPCPSSASGKSFRREQ 371
Db 795 KKYRSCNTHOCPPGSMDFP--REQ 816

RESULT 6
QyVF61 PRELIMINARY; PRT; 1229 AA.
AC QyVF61
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE CG6107 protein.
GN CG6107.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
Ra de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
```

```
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Metalloprotease disintegrin 16 with thrombospondin type I motif.
GN ADAMTS16.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ315734; CAC86015.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1072 AA; 119491 MW; 081EEFF78F47D061 CRC64;

Query Match 32.0%; Score 654; DB 4; Length 1072;
Best Local Similarity 39.5%; Pred. No. 1.8e-56;
Matches 149; Conservative 49; Mismatches 149; Indels 30; Gaps 12;

QY 8 VETLVVADESVMVKFHG-ADLEHYLLTLTATAARLYRHPSILNPINIVVKKVLLLRDRDSG 66
DB 292 VETLVVVDKMKQNHGHENITVYVLTILNMVSAFLKDGKTIGGNINIAIVGLILLEDPG 351
QY 67 PKVTGNAALTLRNCAWKKI-NKVS DKHPEYWDTAILETRQDLGCG--ATTCDPLGMAV 123
DB 352 LVISHADHTUSSFCQWQSLMGKDGTRH----DHAILLTGIDCSKNPEPCDTLGFAP 407
QY 124 GTMCDPKRSCSVIEDGLPSAFTTAHELGHVFNPHDINV-KVCEVFVKGLRANHMSP 182
DB 408 SMCCKSKYRSCNTEDGLGTAFTIAHSGHNFMIHDGEGNCKKSEQ-----NIMSP 462
QY 183 IOIDRANFWSACSAAIITDFLDSGHGDCLLDQPSKPI---SLPDLFGASYTLQQCELA 239
DB 463 AGRNGVFSWSPCSRQYHLKFLSTAQAICLADOP-KPVKEYKYPEKLPGELYDANTOCKW 521
QY 240 FGVGSKPCPY---MOYCTKLWC--TGKAGQMVQCOTRHFHPWADGTSCEGKCLKLGACVE 294
DB 522 FGEKAKLMLDFKDKICKALWCHRGRK-----CETKFWPAEAGTICGDMWCRGGQCVK 576
QY 295 RHNLNKHVRDGSWAKWDPYGCSCRTCCGGVQLARRQCTNPTPANGKGYCEGVRVKYRSCN 354
DB 577 YGDEGPKPTHGHWSWSSWSPCSCRTCCGGVSHRSRLCTNPKPSHGKFCGSTRTLKLCN 636
QY 355 LSPCPSASGKSFREQ 371
DB 637 SOKCPRDS--VDFRAAQ 651

RESULT 8
Q8TE56 PRELIMINARY; PRT; 1095 AA.
AC Q8TE56;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Metalloprotease disintegrin 17, with thrombospondin domains.
GN ADAMTS17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
QY 8 VETLVVADESVMVKFHG-ADLEHYLLTLTATAARLYRHPSILNPINIVVKKVLLLRDRDSG 66
DB 295 VETLVVADKRWVERKHGKNVTYILTVMKVSGLFKDGTGSDINVVVSVLLLEQEPGG 353
```

```
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ315735; CAC86016.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1095 AA; 121099 MW; A5C6540484754D5D CRC64;

Query Match 31.5%; Score 643.5; DB 4; Length 1095;
Best Local Similarity 36.7%; Pred. No. 2.1e-55;
Matches 142; Conservative 64; Mismatches 146; Indels 35; Gaps 12;

QY 8 VETLVVADESVMVKFHGAD-LEHYLLTLTATAARLYRHPSILNPINIVVKKVLLLRDRDSG 66
DB 234 VETLVVADADVMQYHGAQAQRFLITVMVMVYMFQHQSLGKIKINIQVTKLVLLRQPAK 293
QY 67 PKVTGNAALTLRNCAWKKI-----LNKV--SDKHPEYWDTAILETRQDLGCG--GAT 113
DB 294 LSIHGHSRSLSFCHWQNEEYGGARYLGNQVPGKDDPPLVDAAVFVRTDFCVHKDE 353
QY 114 TCDTLGMADVGMCDPKRSCSVIEDGLPSAFTTAHELGHVFNPHDINVVKVCEVFVKGLR 173
DB 354 PCDTVGIAYILGGVCSAKRKCVAEDNGLNLAFTTIAHELGNLGMNHD-----DDHSSCAG 408
QY 174 ANHMSPTLIQIDRANP---WSACSAAIITDFLDSGHGDCLL---DQPSKPISLPEDLP 226
DB 409 RSHIMGEWVK--GRNPSDLSSSCSRDDLENFLEKSKVSTCLLVTDPRSQHTVRLPHKLP 466
QY 227 GASYTLSQCELAFLAGVSGKPCPYNQY--CTKLMCTGKAGQMVQCOTRHFHPWADGTSCEG 284
DB 467 GMHYSANEQCILFGMNATFCRNMEHLMACGLWCL--VEGDTSCTKLDPPLDGTGECGAD 524
QY 285 KCLKGACVERHNLKHVRDGSWAKWDPYGCSCRTCCGGVQLARRQCTNPTPANGKGYCE 344
DB 525 KWRAGECVSKTPIPEH-VGDWSPWGAWSMCSRTCTGARFQRKCDNPPPCPGGTHCP 583
QY 345 GVRVKYRSCNLEPCPSASGKSFREQ 371
DB 584 GASVEHAVCENLPCPKGL--PSFRDQ 608

RESULT 9
Q8TE60 PRELIMINARY; PRT; 1081 AA.
AC Q8TE60;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Metalloprotease disintegrin 21, Last annotation update)
DE ADAMTS18 protein.
GN ADAMTS18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ311903; CAC83612.1; -.
SQ SEQUENCE 1081 AA; 119655 MW; 0438BF645676461E CRC64;

Query Match 31.3%; Score 638.5; DB 4; Length 1081;
Best Local Similarity 38.0%; Pred. No. 6.6e-55;
Matches 143; Conservative 51; Mismatches 153; Indels 29; Gaps 12;

QY 8 VETLVVADESVMVKFHG-ADLEHYLLTLTATAARLYRHPSILNPINIVVKKVLLLRDRDSG 66
DB 295 VETLVVADKRWVERKHGKNVTYILTVMKVSGLFKDGTGSDINVVVSVLLLEQEPGG 353
```

Qy	67	PKVTGNAALTLRNFCAWKKL - NKVSDKHPHYMDTAILETRQDLGG - ATTCDTLGMADV 123
Db	354	LLINHHADQSLNSFCQWSALIGKNGRRH ---DHALLLTGFDICSWKNPCDITLGFAPT 409
Qy	124	GTMCDPKRSQSVIEDDGLPSAFTTAHELGHVFNNPHDNV - KVCEEVFGKLRANHMSPIL 182
Db	410	SGMCKYSRCTINEDTGLGLAFTIAHESGHNFNGMIHDGEGNPCRKAEG -----NINSPIL 464
Qy	183	IQIDRANPWSACSAAIITDFDLSGHGDCLLDQPSK - PTSLPBDLPASVYTSQQCELA 240
Db	465	TGNNGVFSWSQCSQRYLKFLSTPQAGLQDEPKQAGQYKPKLPGQIYDADTKQWQF 524
Qy	241	GVGSKPCP---YMQYCTKLAAC--TGRAKGQWVCQTRHFPWADGTSCEGKGLCLKGACVPR 295
Db	525	GAKAKLSGLGVKDICKSLMCHRVGHR-----CETKMPAAEGTVGGLSMWCRCQGQCVKF 579
Qy	296	HNLNKHVRDGSWAKWDPYGPCSRTCGGGVGLAARQCTNPTPANGKYKCEGVVRKYRSCNL 355
Db	580	GELGPRFIHGOWSAWSKWSRSCRTCGGGYKQFEHCCNNKPKPYGGGIFCPGSSRIYOLCNI 639
Qy	356	EPCSSASGKSFRREQ 371
Db	640	NPC---NENSLDFRAQQ 653

RESULT 10	
Q8WX58	
ID	Q8WX58 PRELIMINARY; PRT; 1223 AA.
AC	OBWX58;
DT	01-MAR-2002 (T-EMBLrel. 20, Created)
DT	01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT	01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE	A disintegrin-like and metalloprotease with thrombospondin type 1 motif 14 precursor.
DE	DE ADAMTS14.
GN	ADAMTS14.
OS	Homo sapiens (Human).
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=21638061; PubMed=11779638;
RA	Bolz H., Ramirez A., von Brederlow B., Kubisch C.;
RT	"Characterization of ADAMTS14, a novel member of the ADAMTS
RT	metalloproteainase family.";
RL	Biochim. Biophys. Acta 152:221-225(2001).
RL	EMBL; AF358666; AAL40229.1; .
DR	InterPro: IPR002870; Pep_M12B_propep.
DR	InterPro: IPR001590; Repolysin.
DR	InterPro: IPR000884; TSP1.
DR	Pfam: PF01562; Pep_M12B_propep; 1.
DR	Pfam: PF01421; Repolysin; 1.
DR	Pfam: PF00090; tsp.1; 4.
DR	SMART; SM00209; TSP1; 4.
DR	PROSITE; PS50215; ADAM_MEPRO; 1.
DR	PROSITE; PS50092; TSP1; 1.
KW	Signal; Incegrin; Protease; Metalloprotease.
FT	SIGNAL 1 22
FT	SEQUENCE 1223 AA; 133871 MW; 363994AC4D92F170F CRC64;

Query Match 30.2%; Score 617.5; DB 4; Length 1223;
Best Local Similarity 35.9%; Pred. No. 9.9e+53;
Matches 134; Conservative 59; Mismatches 155; Indels 25; Gaps 10;

8	QY	VETLVADES	KVPHGAD-LEHYLLTLLATAARLYRHPESIILNPINIVVVVVLLLRDSDG	66
		:	:	:
		:	:	:
		:	:	:
		:	:	:
261	Db	LEVLLVDDSDSVRRP	KEHVYVNTLMNIDEIVHDESLGHHIALVRLINVGVRQL	320
		:	:	:
		:	:	:
		:	:	:
		:	:	:
67	QY	PKV-TGNAALT	LRNFCQWKKLNKYSKHPEDWTAILFTQRDLCGATTCDTLGMADVCT	125
		:	:	:
		:	:	:
		:	:	:
		:	:	:
321	Db	SLIERGPSRLS	LEQVCRWAHSQQRDPSHAEDHDFVFLTRQDF-----GPSGYAPVTG	374

QY	126	MCDPKRSCSVIEDGGLSPSAFTTAHELGHVFNMPHD---	NVKVCEEVFGKLRANHHMMSPTL	182
Db	375	MCHPLRSCLANHEHGFSSAFVIAHGTGHLVMEHDGOGNGCADETSLGS-----	VMAPLV	429
QY	183	IQIDRANPWSACSAAIITDFLDGSHGDCLLDQPSKPI-SLPEDLPGASYTLSQOCELAFG	241	
Db	430	QAATHRFHFWSCSKLELSRPLPS--YDCLLDPDFDPAWPQPELPGINYSWDEQCREFDG	487	
QY	242	VGSKPC-----PYMQYCTKLWCTGKAGQWCVOTRHFHPADWTCGSGKCLKLKACACVERHNL	298	
Db	498	SGYQCLAFRTFECKQWUCS-HPDNPYFCYTKKGGPLDGTETACAPKWCFCGHCIWKSPE	546	
QY	299	NKHRVDGSAWKWDYPGPCSRTCGGGQGLARRQCTNTPANGGKYCEGVRYKRSNCNLEPC	358	
Db	547	QTYQDGGGSWTFKFGSCSRSCGGVRSRSCNNPSPAYGGRPCPLGPMFFQVNCSEEC	606	
QY	359	PSSASGKSFFREQ	371	
Db	607	PGTY--EDFRAQ	617	
RESULT 11				
Q8TE55	PRELIMINARY; PRT; 1223 AA.			
AC	Q8TE55			
DT	01-JUN-2002 (T-EMBLrel. 21, Created)			
DT	01-JUN-2002 (T-EMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (T-EMBLrel. 21, Last annotation update)			
DE	Metalloprotease-disintegrin protease.			
GN	ADAMTS14.			
OS	Homo sapiens (Human).			
OC	EuKaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	TISSUE=FETAL LUNG;			
RX	MEDLINE=21856482; PubMed=11867212;			
RA	Cal S., Obaya A.J., Llanazares M., Garabaya C., Quesada V.,			
RA	Lopez-otin C.;			
RT	"Cloning, expression analysis, and structural characterization of			
RT	seven novel human ADAMTSs, a family of metalloproteinases with			
RL	disintegrin and thrombospondin-1 domains.";			
RL	Gene 283:49-62(2002).			
DR	EMBL; AJ345098; CAC87943.1; -.			
DR	Integrin; Protease.			
KW	SEQUENCE 1223 AA; 133925 MW; D585B6593977ED15 CRC64;			

Query Match 30.2%; Score 617.5; DB 4; Length 1223;
Best Local Similarity 35.9%; Pred. No. 9.9e-53;
Matches 134; Conservative 59; Mismatches 155; Indels 25; Gaps 10

Qy	8	VETLVADESVMVKEPHGAD - LEHYLLTTLTAARLYRHPISILNPINTWVKVLLLRDRDGS	66
Dd	261	IEVLVVDDSVVRHGEKHQNVYLTLMNTVDTEYHDSELGVINIALVRLIMVGVRQSLS	320
Qy	67	PKV-TGNAALTNRNFCQAQKLNKVSDBKPEYWDATLFTRODLGCATTCTDITGMADVGT	125
Dd	321	SLIERGNPSRLEQVCWRHAISOQQRPDSHAHHDHVFLTRQDF-----GPSGYAPVTG	374
Qy	126	MCDPKRSCEVIEDGLPSATTTAHELGHVNMPHD---NVKVEEYFGKLRAHHMMSPTL	182
Dd	375	MCHPLRSCALNHEDGFSFAVIHAETHGVILGMHDGGNGCADETSIGS-----VMAPIV	429
Qy	183	IQIDRANPWSACSAAIITDLDSDGHCLLDLPQSKPI-SLDPDLPCASYTLSQOCELAFG	241
Dd	430	QAAPHRFHWRSCKSLELSRYLPS-YDCLLDLPDPAWPQPPELPPOINTNSMDQCRRFDEG	487
Qy	242	VGSKPC---PYMYCYTKWCTGRAKQMVCOTRHFPWAGTSGGEGKCLKKGACVERHNL	298
Dd	488	SGYQTCLAFTTFBPCQLWCOS-HPDNPNFYCKTKKGPPLDGTCAPOGWKFCKHCITWKSP	546
Qy	299	NKIRVDGSMKWBDPPGPCSRTCGGGQLARRCQTNTPANGGKYCSGVAVKYRSCNLEPC	358

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 11.9712 seconds
(without alignments)
1813.869 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_950
Perfect score: 4043
Sequence: 1 FVSPRYVETLVVADESMVK.....DQCNLRKPKQLDFCVLRPC 738

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2254	55.8	967	US-09-130-491-2	Sequence 2, Appli
2	1911.5	47.3	905	US-09-369-364A-9	Sequence 9, Appli
3	1900	47.0	608	US-09-130-491-13	Sequence 13, Appl
4	1761	43.6	837	US-09-122-126B-2	Sequence 2, Appli
5	1757.5	43.5	1882	US-09-369-364A-13	Sequence 13, Appl
6	1745	43.2	930	US-09-122-126B-15	Sequence 15, Appl
7	1730.5	42.8	874	US-09-369-364A-15	Sequence 15, Appl
8	1726	42.7	930	US-09-369-364A-2	Sequence 2, Appli
9	1607	39.7	551	US-09-130-491-16	Sequence 16, Appl
10	1273	31.5	481	US-09-130-491-8	Sequence 8, Appli
11	1264	31.3	518	US-09-369-364A-22	Sequence 22, Appl
12	1139.5	28.2	1224	US-09-930-872-4	Sequence 4, Appli
13	1084	26.8	997	US-09-369-364A-7	Sequence 7, Appli
14	1060	26.2	1211	US-09-491-522-5	Sequence 5, Appli
15	1033	25.6	1081	US-09-369-364A-17	Sequence 17, Appl
16	1023.5	25.3	1205	US-09-491-522-11	Sequence 11, Appl
17	925.5	22.9	859	US-09-369-364A-5	Sequence 5, Appli
18	846.5	16.0	245	US-09-369-364A-11	Sequence 11, Appl
19	550.5	13.6	525	US-09-369-364A-21	Sequence 21, Appl
20	421.5	10.4	566	US-09-491-522-7	Sequence 7, Appli
21	380.5	9.4	450	US-09-369-364A-19	Sequence 19, Appl
22	332	8.2	491	US-09-930-872-2	Sequence 2, Appli
23	268	6.6	751	US-08-836-443-3	Sequence 3, Appli
24	248.5	6.1	1170	US-08-313-288B-20	Sequence 20, Appl
25	240	5.9	441	US-08-985-526-3	Sequence 3, Appli
26	237	5.9	812	US-09-632-098-4	Sequence 4, Appli
27	227	5.6	802	US-09-632-098-2	Sequence 2, Appli

28	226.5	5.6	855	4	US-09-813-819-2	Sequence 2, Appli
29	226.5	5.6	855	4	US-09-920-048-2	Sequence 2, Appli
30	219	5.4	814	4	US-09-813-819-4	Sequence 4, Appli
31	219	5.4	814	4	US-09-920-048-4	Sequence 4, Appli
32	211.5	5.2	529	2	US-08-836-442-3	Sequence 3, Appli
33	200.5	5.0	788	2	US-08-918-914-4	Sequence 4, Appli
34	199	4.9	616	4	US-09-608-790-1	Sequence 1, Appli
35	199	4.9	1172	1	US-08-313-288B-19	Sequence 19, Appl
36	196.5	4.9	1239	2	US-08-937-931-2	Sequence 2, Appli
37	196.5	4.9	1239	4	US-09-285-502-2	Sequence 2, Appli
38	196.5	4.9	1239	4	US-09-709-126-2	Sequence 2, Appli
39	196.5	4.9	1239	4	US-09-871-385A-2	Sequence 2, Appli
40	193.5	4.8	239	5	PCT-US93-01652-1	Sequence 1, Appli
41	187.5	4.6	613	4	US-09-026-001A-10	Sequence 10, Appl
42	187.5	4.6	621	4	US-09-026-001A-6	Sequence 6, Appli
43	187.5	4.6	621	4	US-09-026-001A-18	Sequence 18, Appl
44	184	4.6	201	4	US-09-411-329C-1	Sequence 1, Appli
45	184	4.6	201	4	US-09-411-335-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-130-491-2
; Sequence 2, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-130-491-2

Query Match	55.8%	Score 2254;	DB 4;	Length 967;
Best Local Similarity	54.0%	Pred. No. 3.3e-185;		
Matches	403;	Conservative 131;	Mismatches 172;	Indels 40; Gaps 11;
QY	1	FVSPRYVETLVVADESMVKFHGADLEHYLLTATAARLYRHPSILNPINIVVVKVLL	60	
Db	253	FVSSHYRVETMLVADQSMAEFHGSLKHYLLTFSVAARLYKHPISIRNSVSLVVKILVI	312	
QY	61	RDRSGPKVTGNAALTNRNCAWOKKLNKVSQKHPEYWDTAILFTRODLCGATTCDTLGM	120	
Db	313	HDEQKPEVTSNAALTNRNCAWOKKLNKVSQKHPEYWDTAILFTRODLCGATTCDTLGM	372	
QY	121	ADVTGMDCKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP	180	
Db	373	ADVTGMDCKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP	432	
QY	181	TLIQIDRANPNWASACNAIITDLDSDHGDCLLDQSPISLPEDLPGASYTLSCQCELAF	240	
Db	433	MLSNLDHSPQSPACNAIITDLDSDHGDCLLDQSPISLPEDLPGASYTLSCQCELAF	492	
QY	241	GVGSKPCP-YMOYCYCKLWCTGKAGQVQCOTRHFPWADGTSCGEGKLCGLKAGCVERINLN	299	
Db	493	GVGSKPCP-YMOYCYCKLWCTGKAGQVQCOTRHFPWADGTSCGEGKLCGLKAGCVERINLN	551	
QY	300	KH---RVDGSAKAWDPYGPCSTCGGGVOLARRQCTNPTPANGKYGCEGVKVRYSNLE	356	
Db	552	KHFDTPFHGSMGMPWGDSCRTCGGVQYTMRECDNFPVKNKGKYGCEGVKVRYSNLE	611	

QY	357	PCSSASGKFRREOCEAFNGYNHSTNRLTLAVAVPKYSGVSPRDKCKLICRANGTYGF	416
Db	612	DCPDN-NGKTFREOCEAHNEFSAKSGSPAVEMTPKYAGVSPKDRCKLICQAKIGYF	670
QY	417	YVLAPKVVDTCLGSPDSTSVCOGKCIKAGCDNLGSKRFRDKCGVCGDGNKSKKVTGL	476
Db	671	FVLQPKVVDTGCPDSTSVCOGQCVKAGCDRIIDSKKFKDKCGVCGGNGSTCKKISGS	730
QY	477	FTKPMHGFVVAIPAAGASSIDIRQYKGLIGDDNYLALKNSQGYLLNGHVFVSAVER	536
Db	731	VTSAPKGYHDIITPTGATNIEVKQNRGSRNNGSFLAKAADGYIILNGDYTLSTLEQ	790
QY	537	DLVVKGSLLRYSGTGRAVESLQASRPILPLTVEVLSVGKWTTPRVYRYSFYLPEKPREDK	596
Db	791	DIMYKGVLYRSGSSALERISFSPLKEPLTTQVLTGVNALRPKIYTYFVKK-----K	846
QY	597	SSHPKDPGRPSVLHNSVLSLNOVEQDDPPARVWAGSWGPCASGSGLOKRAYDCRG	656
Db	847	ES-----FNAIPTFS-----AWTIEWGECSKCELGWQRRLVECRD	883
QY	657	SAGQRTVPACDAH--RPVETQAGE-PCPTWELSAMSPCKSGRQFRRSLKCVGHGG	713
Db	884	INGQ---PASECAKEVPASTRPCADHPQWQLGEWSSCKTCGKGYKRSKLKCLSHDG	940
QY	714	RLARDQCNLHRKQOE-LDFCVLRPC	738
Db	941	GVLSHESCDPLKKPKHFDICFTEMAEC	966
RESULT 2			
US-09-369-364A-9			
; Sequence 9, Application US/09369364A			
; Patent No. 6391610			
; GENERAL INFORMATION:			
; APPLICANT: Apte, Suneel			
; APPLICANT: Hurskainen, Tiina L.			
; APPLICANT: Hirohata, Satoshi			
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases			
; FILE REFERENCE: 26473/4007/10-30-00			
; CURRENT APPLICATION NUMBER: US/09/369,364A			
; CURRENT FILING DATE: 1999-08-06			
; NUMBER OF SEQ ID NOS: 31			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 9			
; LENGTH: 905			
; TYPE: PRT			
; ORGANISM: Mus musculus ADAMTS-8			
US-09-369-364A-9			
Query Match 47.3%; Score 1911.5; DB 4; Length 905;			
Best Local Similarity 50.6%; Pred. No. 8.8e-156;			
Matches 352; Conservative 127; Mismatches 185; Indels 31; Gaps 13;			
QY	1	FVSIPIRYVETLVADESVMVKFHGADLEHYLTLLATARLYRHPHSILNPINIVVKKVLLL	60
Db	229	FYSEARFVETLLVADASMAAFYGTDLQNHILTVMSMAARIYKHPHSIRNSVLNVVKKVLIV	288
QY	61	RDRDSGPKVTGNAALTNRNCAWKLNKVSOKHPEYWDTAILFTFRQDLCG-ATTCDTLG	119
Db	289	EKERGPEVSDNGGLTLRNFCSQWRFNKPSDRHPEHYDTAILFTFRQFCGKGEOCDTLG	348
QY	120	MADVGTMDCKRSCSVIEDDGLPSAFTTAHELGHVFNMPHNDKVKCEEVFGKLRANHMS	179
Db	349	MADVGTICDPKSCSVIKDEGLQAAYTLAHELGHVLSLPHDDSKPCVRLFGPMGYHMAA	408
QY	180	PYLQIDRANPWSACSAIITDFLDSGHGCDLLDQPSKIPISLPEDLPAS--YTLSSQOCE	237
Db	409	PFPIHVNKTLWPSPCSAVYLTLELDDGHGCDLLDAPTSLVPLTGLPGHSLYELDQOQK	468
QY	238	LAFGVSKPCP---YMQYCTKLWCTGKAGQVMVQTRH--FPWADGTSCGEGKCLKGAC	292
Db	469	QIFGPDPRHCPNTSVEDICVLQCARHRSDEPICHKNGSLLLWADGTPCGPGHLCLDGSC	528
US-09-130-491-13			
; Sequence 13, Application US/09130491			
; Patent No. 6416974			
; GENERAL INFORMATION:			
; APPLICANT: Holtzman, Douglas A.			
; APPLICANT: Goodearl, Andrew D.J.			
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83			
; FILE REFERENCE: 09404/041001			
; CURRENT APPLICATION NUMBER: US/09/130,491			
; EARLIER FILING DATE: 1998-08-07			
; EARLIER FILING DATE: 1997-09-05			
; EARLIER APPLICATION NUMBER: US 60/058,108			
; EARLIER FILING DATE: 1997-08-06			
; NUMBER OF SEQ ID NOS: 16			
; SOFTWARE: FastSeq for Windows Version 3.0			
; SEQ ID NO 13			
; LENGTH: 608			
; TYPE: PRT			
; ORGANISM: Mus musculus			
US-09-130-491-13			
Query Match 47.0%; Score 1900; DB 4; Length 608;			
Best Local Similarity 49.3%; Pred. No. 4.9e-155;			
Matches 339; Conservative 114; Mismatches 146; Indels 88; Gaps 8;			
QY	57	VLLLRDRSGPKVTGNAALTNRNCAWKLNKVSOKHPEYWDTAILFTFRQDLCGATTCD	116
Db	4	ILVTHDQKQPEVTSNAALTNRNCAWKLNKVSOKHPEYWDTAILFTFRQDLCGSQTC	63
QY	117	TLMADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHNDKVKCEEVFGKLRANH	176
Db	64	TLMADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHNDKVKCEEVFGKLRANH	176
QY	177	MMSFTLIQIDRANPWSACSAIITDFLDSGHGCDLLDQPSKIPISLPEDLPAS--YTLSSQO	236
Db	124	MMSFTLIQIDRANPWSACSAIITDFLDSGHGCDLLDQPSKIPISLPEDLPAS--YTLSSQO	236
QY	237	ELAGVSKPCP-YMQYCTKLWCTGKAGQVMVQTRH--FPWADGTSCGEGKCLKGACVER	295
Db	184	QTFEGEDSKHCPDAASTCTSLWCTGSGGVLCVQCTKHFHWADGTSCGEGKCLKGACVER	243

Db	453	FGPDSRHCPQLPPCCAALWCSGHLNHAMCQTKHSPWADGTPCQPAQACMGGRCLHMDOL	512
QY	299	NKHYVD--GSWAKWDYPCPSRTCGGGVOLARQCTNPPTPANGKGYCEGVVRYKJSCNLE	356
Db	513	QDFNTPQAGGPGWPGWDCSRTCGGGVQFSRRDCTRPVRPNGKGYCEGRTRFSCNTE	572
QY	357	PCPSSASCKSFREQCEAFNGYNHSTNRLLT---AAVAVPKYSGVSPDRCKKLICRANGT	413
Db	573	DCP-TCSALTFREQCAA---YNHRTDLKSPFGPMDWVPRYTVAPQDOCKLTCCOARL	628
QY	414	GYFYVLAPKVVDGTLCSPDSTSCVCGVKCIKAGCDGNLGSKKRPDKCGVCGDNKSKKV	473
Db	629	GYYYVLEPRVVDGTPCSPDSSVCVCGRCHAGCDRIIGSKKKFDKCMVCGGSGGCSKO	688
QY	474	TGLETKPMHGNFYVVAIPAGASSIDIRQRYGKGLIGDDNYLALKNOSGKYLLNGHFWSA	533
Db	689	SGSFRKFRGYNNVVTIPAGATHILVRQONGP---HRSYILALALPDGVSALANGEXTLMP	746
QY	534	VERDLVVRGSL-LAYSGTGTAVESLQASRPILLEPTLVESLVGKMTTPRVRYSFYLPKEP	592
Db	747	SPTDWWPLGASVLSYGATASELISGHGFLAQPLTLQVLVANGNPQDTRLRYSEFFVRPT	806
QY	593	REDKSGHPKD	602
Db	807	PSTPRTPQD	816

RESULT 5
US-09-369-364A-13
; Sequence 13, Application US/09369364A
; Patent No. 6391610

```

1  APPLICANT: Hurskainen, Tiina L.
2  APPLICANT: Hironaka, Satoshi
3  TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
4  FILE REFERENCE: 26473/4007/10-30-00
5  CURRENT APPLICATION NUMBER: US/09/369,364A
6  CURRENT FILING DATE: 1999-08-06
7  NUMBER OF SEQ ID NOS: 31
8  SOFTWARE: PatentIn Ver. 2.1
9  SEQ ID NO 13
10 LENGTH: 1882
11 TYPE: PRT
12 ORGANISM: Homo sapiens ADAMTS-9
13 FEATURE:
14 NAME/KEY: MOD_RES
15 LOCATION: (468)
16 OTHER INFORMATION: Xaa = C
17 NAME/KEY: MOD_RES
18 LOCATION: (521)
19 OTHER INFORMATION: Xaa = Y
20 US-09-369-364A-13

```

Qy	1	FVSIPIRYVTLLVVADESVMVKEHGADLBHYLLTLLATAARLYRHPISILNPINIVVVKVLLL	60
Db	236	FLSYPRFVELVVAADNRMVSYHNGLOHYLLTLLMSIVASIKYDPSIGNLINVIVNLIVI	295
Qy	61	RDRSDGPKYTGNAALTLRNFCAWOKLKNYSKDHPEYWFDAILETFRODLCA--TTCDTLG	119
Db	296	HNEQDGSFSFNQYTLTKNFCQWHSNPSGGIIH---DTAVLLTRQDICRAHDKCDTLG	351
Qy	120	MADVGTMCDDPKRSCSVIEDDDGLPSAFTIAHELGHVFNMFPMNVKCYEEVFGKLRANHMS	179
Db	352	LAEGLTICDPYRSCISSEDSGLSTAFIAHELGHVFNMFPMNDONNCKEE--GVKSPQHVA	410
Qy	180	PTLIQIDRANPWSACSAALITDFDLSHGDCLLDQP--SRPISLPEDLPGASYTILSQOCEL	238
Db	411	PTLNFTNYTNPMWMSKSRKITYTEFLDTQYGCLELNEPSRPYLPQVQLPGYLLNVKQXEL	470

Qy	239	AFVGSKPCPYMOYCTKLWC---	TGKAKGMVQCOTRHFPWADTSCGEGKCLKGACVER	295
Db	471	IFGSGQVCPYMQCRLWNNVYVHGK---	CRQTHTPWADTECEPKGKXGFCVPK	527
Qy	296	HNLNKHVDGSKWAKWDPYPCSRCTCGGCVOLARRQCTNPT	PANGGKYCEGVRYKYRSCNL	355
Db	528	-EMDVPYDGSWSWSFGCTSRCTCGGKIATRECNPEPKNGKYCYVGRMKFKSCNT	586	
Qy	356	EPCPSASGSKSFREOCEAFNGYNHSTNRLTAVAVWPYKSYGVSPPDKCKLICRANGTGY	415	
Db	587	EPCLKQK--RDFRDEQCAHEFDGKHFNINGLLPNVRWVPYKSYGILMKDRCKLFCRVAGNTA	644	
Qy	416	FYVLAPKVDGTLCSPDSTSVYQKCKIKAGCDNGLGSKKRFDKCGVCGGDNCKSKKVTG	475	
Db	645	YVOLRDVYIDGTPCGDNDICVQGLCRQAGCDHVLNKSARRDKCGVCGGDNSSCKTVAG	704	
Qy	476	LFTKPMHGYNFVAIPAGASSIDIRQYKGLIGDDNYLALKNSQKYLINGHFVVSARE	535	
Db	705	TFTNVHYGYNVVRIPAGATNIDVRQHSFSGTDDDDNYLALSSKGEFLNGNFVVTMAK	764	
Qy	536	RLVVKSLRLYSGTGTAVESLQASRILEPLTVEVLVSGKMTPPRVRYSYFLPKPEPRD	595	
Db	765	REIRIGNAVVEYSGSETAVERINSTRIEQLLQVLSVGLKYNPDVRYSFNPIEDKQP	824	
Qy	596	K---SSHPKDPGRP-----	SVLHNSVLSLSNQVEQDPRPP-----	628
Db	825	QFYWNH-----	GPWQACSKPCQGERKKRLVCTRESQDQTVS---DQCRDLRPPQGHTE	876
Qy	629	-----ARWAGSWGPCASCGSLQKRAVDCR-----	GSAGQRTVPACDAARHPVET	675
Db	877	PCGTGCDLRHWHVASRSECSAQCGLYRTLDIYCAKYSRLDGKTEKVDGFCSSHHPKPSNR	936	
Qy	676	QACGECPT--WELSAWSPSCSGRGFORRLKCVGHHGRLARDQCNLHRKPOELDFC	733	
Db	937	EKCSGECTGWRYSANTSCSKDGGTQRRRAICVNRDVLDDSKCT-HQEKVTIQR	995	
Qy	734	VLRPC 738		
Db	996	SEFPC 1000		
RESULT 6				
US-09-122-126B-15				
; Sequence 15, Application US/09122126B				
; Patent No. 6451575				
; GENERAL INFORMATION:				
; APPLICANT: Bristol-Myers Squibb Company				
; TITLE OF INVENTION: AGGREGAN DEGRADING METALLO PROTEASES				
; FILE REFERENCE: DM6909				
; CURRENT APPLICATION NUMBER: US/09/122,126B				
; CURRENT FILING DATE: 1998-07-24				
; NUMBER OF SEQ ID NOS: 21				
; SOFTWARE: PatentIn version 3.0				
; SEQ ID NO 15				
; LENGTH: 930				
; TYPE: PRT				
; ORGANISM: Homo sapiens				
US-09-122-126B-15				
Query Match 43.2%; Score 1745; DB 4; Length 930;				
Best Local Similarity 48.7%; Pred. No. 1.9e-141;				
Matches 335; Conservative 98; Mismatches 219; Indels 36; Gaps 10;				
Qy	2	VSIPRYVELTVVADESVMKFGADLEHYLLTLTAARLYRHPSILNPINIVVVKVLLR	61	
Db	263	ISRARQVELLVADASMARLYGRGLQHYLLTLASIANRLYSHASTENHRLAVKVVVLG	322	
Qy	62	DRDSGPKVTGNAAITLRFNCAWKKLNKVSXKHPYWDATILFTQDLCGATCTDLGNA	121	
Db	323	DKDKSLEYSKNAATLKNFKCKWQHQNQLGDHDEEYDAAILFTREDLCGHSCDTLQMA	382	
Qy	122	DVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKVCBEVFGKLRANHMSPT	181	

[illegible][illegible]

Matches	327;	Conservative	105;	Mismatches	218;
				Indels	40;
				Gaps	9;

```
Db 1 CASLNGVSGDHLWASMLSLDHSQWSPSCSAIYMTVTSFLDNGHGECMLDKPQNKLPSD 60
QY 225 LPGAASYLSQOCELARGVSKPCP-YMQYCTKLWCTGKAKGOMVQCTREFPWADGTSCE 283
Db 61 LPTGLDANQOQCTTGEESKHCPCDAASTCTTLMCTGTSGGLVVCOTKFFPWADGTSCE 120
QY 284 GKLCAGACVYRHNLNK---RYDGSWAKWDPYGPCSRTCTGGGVQLARRQCTNPTPANG 340
Db 121 GKVCVSKCYNKTDK-KHFATPVHGSWGPWPGDCSRTCTGGGVQVYTMRECDNPVPKNG 179
QY 341 KYCEGVVYKRSNLEPCPSSASGKSFREOCEAFNGYNHSTNRLTLAVAWVPKYSGVSP 400
Db 180 KYCEGRVYRSCNIEDCPDN-NGKTFREOCEAHNEFSKASFGNEPTVEWTPKYAGVSP 238
QY 401 RDKCKLICRANGTFYFVLAPKVVDTGLSPDSTSCVCGKCIKAGCDNGLSKKRFDC 460
Db 239 KDKCKLCEAKGIGYFVLQPKVVDTGPCSPDSTSCVCGQCVKACDRIIDSKKFKDC 298
QY 461 GVCGGDNKSKKVTGLTKPMHGYNFVAIPAGASSIDIRQRYKGLIGDDNYLALKNQSO 520
Db 299 GVCGGNGSTCKKMSGIYVTSRPGYHDVTIPAGATNIEVKHRNORNGSRNGSFLAIRAD 358
QY 521 GKYLNGHFVSAVERDLVVGSLRYSGTGAVESIOASRPILPLEPTVEVLVSGKWTTP 580
Db 359 GTYLNGFTLSTLEQDLTKVILRYSGSAALERIRSFSPKLEPLTIQVLVGHALRP 418
QY 581 RVRYSFYLPKPREDKSSHPKDPGRPSVLHNSVLSLNOVEQPDPRPPARWAGSWGPCS 640
Db 419 KIKTFYPMKKKTES-----FNAIPTFS-----EWVIEENGEC 451
QY 641 ASCSGYLOKRAVDCRGAGORTVPADAAH--RPVETOACGE-PCPTWELSAWSPCKSC 697
Db 452 KTCGSGWQRRVVOCRDINGH---PASECAKEVKPASTRPCADLPCPHWQVGDWSPCKSC 508
QY 698 GRGQFRESLKVCGHGGHLLARDONLHRKPQE-LDFCVLRPC 738
Db 509 GKGYKRTKLKCVSHGGVLSNESCDPLKPKHYIDFTLTQC 550

RESULT 10
US-09-130-491-8
; Sequence 8, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; EARLIER FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 481
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-130-491-8

Query Match 31.5%; Score 1273; DB 4; Length 481;
Best Local Similarity 47.0%; Pred. No. 3e-101;
Matches 236; Conservative 87; Mismatches 141; Indels 38; Gaps 12;

QY 190 PWSACSAIITDFDLSHGDCDLLDQPSKPISLPDLPGASYLSQOCELAFGVGSKPCP- 248
Db 7 PWSPCSAVYLTLLDGDHGDCLLD-----DGHSTLYELDQOQKQIFGDFDRCPN 56
QY 249 --YMQYCTKLWCTGKAKGOMVQCTRH--FPWADGTSCEGKCLKLGKACVERHNLN--KHR 302
Db 57 TSVEDICVQLWCRHRSDEPICHTKNLSLLWADGTPCGPHLCLDGLDSCVLRVEENPKAV 116
```

```
QY 303 VDGSWAKWDPYGPCSRTCTGGGVQLARRQCTNPTPANGKYCEGVVYKRSNLEPCPSSA 362
Db 117 VDGDWGPWGPWGQCSRTCTGGGIQFSNRECDNPAPQNGRFLGERVKYQSKTEEC--P 174
QY 363 SGKSFREOCEAFNGYNHSTNRLTLAVAWVPKYSGVSPDKCKLICRANGTGYFYVLAPK 422
Db 175 NGKSFREOQCCYKAYNH-TLDNGFNLOWVPKYSGVSPDRCKLFCRARGSEKVFETK 233
QY 423 VVDGTLCPDSTSCVCGKCIKAGCDNGLSKKRFDCGVCGGDNKSKKVTGLTKPMH 482
Db 234 VIDGTLCPDILAICVRQCVKAGCDHVNNSPKLDCGVCGGGTACRKYSGSFTPEY 293
QY 483 GYNFVAIPAGASSIDIRQRYKGLIGDDNYLALKNQSKYLLNGHFVVSVERDLVVGK 542
Db 294 GYNDVTVTIPAGATNIDVKQRSHPGVQNDGYLAKLTANGQYLLNGLNLAISAIEDILMK 353
QY 543 SLLRYSGTGTAVESIOASRPILPLEPTVEVLV-S-GKMTPPRVRYSFYLPKPREDKSSHPK 601
Db 354 TLLKYSGSMATLERLQSFALPELTVQLLTVSGEVFPKPKYTFVFPND--TDFNVQSS 411
QY 602 DPRGPSVLHNSVLSLNOVEQPDPRPPARWAGSWGPCSASCSGLQKRAVDCRGSAQOR 661
Db 412 KERAST--NLIQSL-----PYAEVLGDWSECPSTCGGQWQRTVECDRPPSQA 458
QY 662 TVPADCAAHRPVETQACG-EPC 682
Db 459 S-DTCDCAKPEDAKPCGSOQC 479

RESULT 11
US-09-369-364A-22
; Sequence 22, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-5
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (99)
; OTHER INFORMATION: Xaa = M
US-09-369-364A-22

Query Match 31.3%; Score 1264; DB 4; Length 518;
Best Local Similarity 45.4%; Pred. No. 2e-100;
Matches 244; Conservative 73; Mismatches 185; Indels 36; Gaps 10;
```

```
QY 152 GHVFNNPHDNVYCEVEFGKLRANHMMSPTLIQIDRANPWSACSAIITDFDLSHGDC 211
Db 1 GHLGLSHDSSKFCETFGSTEDKRLMSSILTSIDASKPWSKTSATITEFLDDHGNC 60
QY 212 LDQPSKPISLPDLPGASYLSQOCELAFGVGSKPCPYMYCTKLWCTGKAKGOMVQCTR 271
Db 61 LDLPRKQILGPPELPQQTVDATQCCNLTFGPDYSVPCGXDXVCARLCAVAVROGVCLTK 120
QY 272 HFPWADGTSCEGKCLCKACACVER-----HNLNKRVDSGWAKWDPYGPCSRTCTGGVQL 326
Db 121 KLPAVEGTGCGGRICLQCKVDKTKKKYISTSSH---GNWGSWGSWCSGCSGGGVQF 177
QY 327 ARROCTNPTPANGKYCEGVVYKRSNLEPCPSSASGKSFREOCEAFNGYNHSTNRL 386
Db 178 AYHCNPNAPRNGRYCTCKRAIYHSCSLMPCP--PNGKSFREOCEAKNGYQSDAKGVK 235
```

Db	637	SQCKPRDS--VDFRAAQCAEHNRRRRRGHYK-----WKP-YTQVEDQDLCKLYCI	684
Qy	410	ANETGYFVLAPKVDGTCLSPDSTVCVQGCRCIKAGCDNGLSGKKRFDKGVGGDNKS	469
Db	685	AEQDFDFEFLSNVKDGTPCSDRSRNVCTDGCERVGCNDVLGSDAVEDVGCVCNGNSA	744
Qy	470	CKKVTLGFTKPMH--GYNFVAIPACAGASSIDTRQRCYKGLIGDDNYLALKNSQGYLLNG	527
Db	745	CTTHRGLYTKHHHTNOYYHMVTIPSGARSIRYEMNV-----STSYISVRNALRRYYLNG	799
Qy	528	HFVVSAYERDLVVKGSLRLYSGTGTAVESLQASRPITLEPLTVEVLVSVGKMTPPRVKXSFY	587
Db	800	HWTVDMPGR-YKPSGTTFDYRSYNENPLIATGPNETLIVELLFOGR--NPGVAMEYS	856
Qy	588	LPKEPREDKSSHKDPGRPS-----VLHNSVLSLNOVE-----QP	623
Db	857	MPR-----LGTEKQPPAQPSYTWAIVRSECVSGGGQMTVREGCYRDLKFQVNMSCNP	911
Qy	624	DDRP-----PARVAGSWGPCASCGSLGLOKRAVDC--RGSAGORTVPA--CD	667
Db	912	KTRPVTGLVCKYVSACPPSWSVGNWSACSRTCGGAQSRPVQCTRHHVYDSEPVASLCP	971
Qy	668	AAHRPVETQAC-GEPC-PTWELSAWSPCKSGKGRGFRQRSLKCVGHG---GRLLARDQC	721
Db	972	QP-APSSRQACNSQSPPAWSAGPWAECSHTCGKWKREKAVACKTNP SARAQLLPDAVC	1030
Qy	722	NLHRKKPOELDFCVLRPC	738
Db	1031	TSEPPKPRMHEACLLQRC	1047

```

RESULT 13
US-09-369-364A-7
; Sequence 7, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
US-09-369-364A-7

```

```

; LENGTH: 99/
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
US-09-369-364A-7

Query Match          26.8%; Score 1084; DB 4; Length 997;
Best Local Similarity 37.2%; Pred. No. 1.5e-84;
Matches 278; Conservative 90; Mismatches 303; Indels 76; Gaps

```

Query Match	26.8%;	Score 1084;	DB 4;	Length 997;
Best Local Similarity	37.2%;	Pred. No. 1.5e-84;		
Matches 278; Conservative 90; Mismatches 303; Indels 76; Gaps				
QY	2	VSTPRVETLVWADSESMVKFHG-ADLEHVLTLTLLATAARLHRHPHSILPINIVVKVLLL	60	
Db	238	VSKEKCETLVWADAAMVETHGQPVESVLTIMNMVAGLFHDPSIGNPIHTIVRLVLL	297	
QY	61	RDRDSGPKVTGNAAALTRIFCAWQKLNKYSDKHPEYMDTAILETRDQLCGATT--CDTL	118	
Db	298	EDEEDLKITHHADNTLKSFCKWKQSINMKGDAPHLPHDDTAILLTRKDLCAMNRPCELT	357	
QY	119	GMAADVGTMCDPKRSCSVIEDGLPISAFTHAELGHVFNNPHD-NVKVCEEVFGKLRAHHM	177	
Db	358	GLSHVAGMCQPHRSCSINETGLPTAFTVAHELGHFSFIQHDDSGNDCEPY-GK-RPFI	414	
QY	178	MSPTLIQIDRAPNWSACSAAITDFUDSGHGCLLDQSPK-TISPEDLPCASTLTQQC	236	
Db	415	MSPOLLYDAAPLTWSCSRQYITRFIDRWGWLCLDDPPAKDIIDPSPVPGLVDVSHQC	474	
QY	237	ELAFGVGSKPCPYW-QYCTKLWCTGRAKGMVCQTFRHFPPADWTSCGGKLCCLKAGVER	295	

GenCore version 5.1.4_p5_4578
 Copyright (c) 1993 - 2003 Compugen Ltd.
 OM protein - protein search, using sw model
 Run on: April 29, 2003, 17:12:21 ; Search time 16.4049 Seconds
 (without alignments)
 3604.758 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_950
 Perfect score: 4043
 Sequence: 1 FVSIPRYVELLVVADESWK.....DQCNLHRKPQELDFCVLRPC 738

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5
 Searched: 301932 seqs, 80129803 residues
 Total number of hits satisfying chosen parameters: 301932
 Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4043	100.0	950	10	US-09-965-631-4
2	3136.5	77.6	823	9	US-10-163-316-2
3	2260	55.9	727	9	US-10-097-597-1
4	2260	55.9	727	9	US-10-097-580-1
5	2260	55.9	727	10	US-09-445-023A-1
6	2259	55.9	727	9	US-10-097-597-12
7	2259	55.9	727	9	US-10-097-580-12
8	2259	55.9	727	10	US-09-445-023A-12
9	2257	55.8	950	10	US-09-321-987B-4
10	2257	55.8	968	9	US-10-163-316-7
11	2254	55.8	967	12	US-10-105-929-2
12	1911.5	47.3	905	10	US-09-918-171A-9
13	1901	47.0	608	10	US-09-803-589-2
14	1900	47.0	608	10	US-09-803-589-8
15	1900	47.0	608	12	US-10-105-929-13
16	1779	44.0	1829	10	US-09-972-467-2
17	1760	43.5	837	9	US-10-174-590-352
18	1760	43.5	837	9	US-10-176-758-352
19	1760	43.5	837	9	US-10-175-737-352

20	1760	43.5	837	9	US-10-173-706-352	Sequence 352, App
21	1760	43.5	837	9	US-10-175-738-352	Sequence 352, App
22	1760	43.5	837	9	US-10-175-752-352	Sequence 352, App
23	1760	43.5	837	9	US-10-176-482-352	Sequence 352, App
24	1760	43.5	837	9	US-10-176-757-352	Sequence 352, App
25	1760	43.5	837	9	US-10-176-913-352	Sequence 352, App
26	1760	43.5	837	9	US-10-180-552-352	Sequence 352, App
27	1760	43.5	837	9	US-10-180-557-352	Sequence 352, App
28	1760	43.5	837	9	US-10-173-700-352	Sequence 352, App
29	1760	43.5	837	9	US-10-174-572-352	Sequence 352, App
30	1760	43.5	837	9	US-10-174-579-352	Sequence 352, App
31	1760	43.5	837	9	US-10-174-582-352	Sequence 352, App
32	1760	43.5	837	9	US-10-174-588-352	Sequence 352, App
33	1760	43.5	837	9	US-10-175-739-352	Sequence 352, App
34	1760	43.5	837	9	US-10-175-740-352	Sequence 352, App
35	1760	43.5	837	9	US-10-175-743-352	Sequence 352, App
36	1760	43.5	837	9	US-10-176-488-352	Sequence 352, App
37	1760	43.5	837	9	US-10-176-492-352	Sequence 352, App
38	1760	43.5	837	9	US-10-176-747-352	Sequence 352, App
39	1760	43.5	837	9	US-10-176-750-352	Sequence 352, App
40	1760	43.5	837	9	US-10-176-985-352	Sequence 352, App
41	1760	43.5	837	9	US-10-176-987-352	Sequence 352, App
42	1760	43.5	837	9	US-10-176-991-352	Sequence 352, App
43	1760	43.5	837	9	US-10-176-992-352	Sequence 352, App
44	1760	43.5	837	9	US-10-176-993-352	Sequence 352, App
45	1760	43.5	837	9	US-10-184-658-352	Sequence 352, App

ALIGNMENTS

RESULT 1
 US-09-965-631-4
 ; Sequence 4, Application US/09965631
 ; Patent No. US20020115842A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Friddle, Carl Johan
 ; APPLICANT: Hilbun, Erin
 ; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Enco
 ; FILE REFERENCE: LEX-0241-USA
 ; CURRENT APPLICATION NUMBER: US/09/965,631
 ; CURRENT FILING DATE: 2001-09-27
 ; PRIOR APPLICATION NUMBER: US 60/236,689
 ; PRIOR FILING DATE: 2000-09-29
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 950
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 US-09-965-631-4

Query Match 100.0%; Score 4043; DB 10; Length 950;
 Best Local Similarity 100.0%; Pred. No. 2.4e-287;
 Matches 738; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	FVSIPRYVELLVVADESWKFGADLEHYLLTLATAARLYRHPHSILNPINIVVKVLLL	60
DB	213	FVSIPRYVELLVVADESWKFGADLEHYLLTLATAARLYRHPHSILNPINIVVKVLLL	272
QY	61	RDRDSGPKVTGNAALTNRNFCWOKKLNKVSQKHPEYWDTAILFTRODLGGATTCDTLGM	120
DB	273	RDRDSGPKVTGNAALTNRNFCWOKKLNKVSQKHPEYWDTAILFTRODLGGATTCDTLGM	332
QY	121	ADVGTMCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP	180
DB	333	ADVGTMCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP	392
QY	181	TLIQIDRANPNASAAIITFDLSGHGDCLLDQSPKIPSLPDLPGASYTLSSQCEELAF	240
DB	393	TLIQIDRANPNASAAIITFDLSGHGDCLLDQSPKIPSLPDLPGASYTLSSQCEELAF	452
QY	241	GVGSKPCPYMOYCTKLWCTGKAGQMVQTRHFFWADTSCGEGKCLKGACVERHNLN	300

Db	453	GVSGKCPYMQYCTKLWCTGKAGQVMVQCTRHPWADGTSCEGKGLCLKGACVERHNLNK	512
Qy	301	HRVDGSWAKWDPPGCSRSCGGGVQIARQCTNPPTPANGKCYCEGVRKYRSCNLEPCPS	360
Db	513	HRVDGSWAKWDPPGCSRSCGGGVQIARQCTNPPTPANGKCYCEGVRKYRSCNLEPCPS	572
Qy	361	SASGKSPRESCQAFNGYNHSTNRLTLAVAWPKYSGVSPROCKLICRANGTYFYVLA	420
Db	573	SASGKSPRESCQAFNGYNHSTNRLTLAVAWPKYSGVSPROCKLICRANGTYFYVLA	632
Qy	421	PKVVDGTLCSPDSTSVCGKCIKAGCDGNLGSKKRFDKCGVGGDNKSCKKVTGLTFKP	480
Db	633	PKVVDGTLCSPDSTSVCGKCIKAGCDGNLGSKKRFDKCGVGGDNKSCKKVTGLTFKP	692
Qy	481	MHGYNFVVAIPAGASSIDIRQCYKGLIGDDNYLALKNSQGYLLNGHPVVSVERDLVV	540
Db	693	MHGYNFVVAIPAGASSIDIRQCYKGLIGDDNYLALKNSQGYLLNGHPVVSVERDLVV	752
Qy	541	KGSLRLRSGTGTAVESLQASRPILPTLVEVLSVGKMTPPRVRYGFYLPKEPREDKSSH	600
Db	753	KGSLRLRSGTGTAVESLQASRPILPTLVEVLSVGKMTPPRVRYGFYLPKEPREDKSSH	812
Qy	601	KDPRGPSVLHNSVLNSQYEQDDPRPARWAGSWGPCSSCGSGLOKRAVDCRGSAGQ	660
Db	813	KDPRGPSVLHNSVLNSQYEQDDPRPARWAGSWGPCSSCGSGLOKRAVDCRGSAGQ	872
Qy	661	RTVPACDAAHRPVBETQACGPECPWTWELSAWSPCSKSGRFFRRSLKCYGHGGRLLARQD	720
Db	873	RTVPACDAAHRPVBETQACGPECPWTWELSAWSPCSKSGRFFRRSLKCYGHGGRLLARQD	932
Qy	721	CNLHRKKPOELDFCVLRPC	738
Db	933	CNLHRKKPOELDFCVLRPC	950

```

RESULT 2
US-10-163-316-2
; Sequence 2, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MPI01-025PIRNM
; CURRENT APPLICATION NUMBER: US/10/163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: fastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-316-2

```

181	Qy	TLIQIDRANPWSACSAAIITDFDLSGHDCLLDQPKPSLPEDLPGASYTILSQOCELAF	241
393	Db	TLIQIDRANPWSACSAAIITDFDLSGHDCLLDQPKPSLPEDLPGASYTILSQOCELAF	452
241	Qy	GVGSKPCPYMOYCTKLWCTGKAKQMWQCTRHPFWADGTSCEGSKLCLKGACVERHNLNK	300
453	Db	GVGSKPCPYMOYCTKLWCTGKAKQMWQCTRHPFWADGTSCEGSKLCLKGACVERHNLNK	512
301	Qy	HR-----VDGSWAKWDPYGPCSRTCGGVQVLARROCTNPTPA	337
513	Db	HRPPTDII SPQKLLRLPNGLHTTQVDGSWAKWDPYGPCSRTCGGVQVLARROCTNPTPA	572
338	Qy	NGGYCEGVRYVKYRSCNLEPCPSSASGKSFPREOCEAFNGYNHSTNRLTLAVAWVPKYSG	397
573	Db	NGGYCEGVRYKYRSCNLEPCPSSASGKSFPREOCEAFNGYNHSTNRLTLAVAWVPKYSG	632
398	Qy	VSPDKCKLICRANGTGYFVYLAPKVVDGTLCPSPDSTSVCVQCKCIKAGCDGNLGSKKRF	457
633	Db	VSPDKCKLICRANGTGYFVYLAPKVVDGTLCPSPDSTSVCVQCKCIKAGCDGNLGSKKRF	692
458	Qy	DKCGVCGGDNNKSKKVTGLFTKPMHGYNFVVAIPAGASSDIDIRQGYKGLIGDNNYLAK	517
693	Db	DKCGVCGGDNNKSKKVTGLFTKPMHGYNFVVAIPAGASSDIDIRQGYKGLIGDNNYLAK	752
518	Qy	NSQGYLLNGHFVYSAYVERDLVVKGSLLRYSGTGTAVESLQASRPILPTVEVLVSVGKM	577
753	Db	NSQGYLLNGHFVYSAYVERDLVVKGSLLRYSGTGTAVESLQASRPILPTVEVLVSVGKM	812
578	Qy	TPPR 581	
813	Db	TPPR 816	

```

RESULT 3
US-10-097-597-1
; Sequence 1, Application US/10097597
; Publication No. US2003002232A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Ei-ji
; APPLICANT: Hakoizaki, Michinori
; APPLICANT: Ishioka, Kelko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: pharmaceutical
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-597-1

```

```
Db 73 HDQKGEVTSNAALTLRNFNCWKQKOHNPSPDRDAEHYDTAILFTRODLGSGQTCDTLGM 132
QY 121 ADVTGTCWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNDVKVCEEVFGKLRANHMSP 180
Db 133 ADVTGTCWCDPKRSCSVIEDDGLQAAFTTAHELGHVFNPHDNDVKVCEEVFGKLRANHMSP 192
QY 181 TLIOIDRANPWSACSAALITDLDGSHGDCLLDQPSKPISLPEDLPFGASYTLSQOCELAF 240
Db 193 MSLNLDHSPWSPCSAYMITSLDNGHGECLMDKPNQIQLPGDLPTGLTDANRQCQTF 252
QY 241 GVCSKPCP-YMQYCTKLWCTGAKQWCVQTRHFPWADTSGEGKCLKLGACVHERNLN 299
Db 253 GEDSKHCPDAASTCTLWCTGSGVLVQTKHFPWADTSGEGKCLKLGACVHERNLN 311
QY 300 KH---RVDGSAKWDYPGCSRTCGGVQVQARROCTNPTPANGKYCEGVRYKRSCLNLE 356
Db 312 KHFDTPFHSGWPGWPGWDCSRTCGGVQVQARROCTNPTPANGKYCEGVRYKRSCLNLE 371
QY 357 PCPSSASGKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYG 416
Db 372 DCPDN-NGKTFREOCEAHNEFSKASFGSGPAVEWIPKYAGVSPKDRCKLICQAKGIGYF 430
QY 417 YVLAPKVVDTGLCSPDSTSVQVQKICAGCDNGLSKRFPDKGVCGGDNKSKRVKVTGL 476
Db 431 FVLQPKVVDTGLCSPDSTSVQVQKICAGCDRIIDSKKFKDKGVCGGNGSTCKKISGS 490
QY 477 FTKPMHGYNFVAIPAGASSIDIRQRYKGLIGDDNYLALKNQSGKYLINGHVFVSAVER 536
Db 491 VTSKPGYHDVTIPTGATNIEVKQRNQRNNGSFLAKAADGTIYLNGDYTLSTLEQ 550
QY 537 DLVVKGSLRYSGTGTAVESLQASRPILPLETVESLVSGKMTPPRVRYSFYLPKPREDEK 596
Db 551 DIMYKGVWLRYSGSSAALERIRSFPLKEPLTIQVLVGNALRPKIYTFVKK- ---K 606
QY 597 SSHPKDPRGSPVLHNSVLSNQVEQDPRPARVAGSWGPCSASGSGLQKRAVDCRG 656
Db 607 ES-----FNAIPTFS-----AWVIEWEGCSKSCELGWQRRLVECRD 643
QY 657 SAGORTVPACDAAH--RPVETOACGE-PCPTWELSAWSPCSKSCGRGFORRLSKCVHGG 713
Db 644 INQO---PASECAKEVKAPEASTRPCADHPCPQWOLGEWSSCSKTCGKYKRLKCLSHDG 700
QY 714 RLLARQDCNLRKPKQE-LDFCVLRPC 738
Db 701 GVLSHESCDPLKPKHFIDFCTLTQC 726
```

RESULT 4

```
US-10-097-580-1
; Sequence 1, Application US/10097580
; Publication No. US20030032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiji
; APPLICANT: Hakozaiki, Michinori
; APPLICANT: Ishioke, Keiko
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; TITLE OF INVENTION: Composition and method of immunologically analyzing human ADAMTS
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-10-097-580-1
```

```
Query Match 55.9%; Score 2260; DB 9; Length 727;
Best Local Similarity 54.3%; Pred. No. 3.8e-157;
Matches 405; Conservative 129; Mismatches 172; Indels 40; Gaps 11;
```

```
QY 1 FVSPRVVETLVVADESVMVKFHGADLHYLLTLATLAARLYRHPHSILNPINIVVVKVLL 60
Db 13 FVSPRVVETMLVADQSMAEFHSGLKHYLLTLEFSVAARLYKHPHSIRNSVSLVVKVILVI 72
QY 61 RDRSDGKPVKTGNAALTLRNFNCWKQKOHNPSPDRDAEHYDTAILFTRODLGSGQTCDTLGM 120
Db 73 HDEQKGEVTSNAALTLRNFNCWKQKOHNPSPDRDAEHYDTAILFTRODLGSGQTCDTLGM 132
QY 121 ADVTGTCWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNDVKVCEEVFGKLRANHMSP 180
Db 133 ADVTGTCWCDPKRSCSVIEDDGLQAAFTTAHELGHVFNPHDNDVKVCEEVFGKLRANHMSP 192
QY 181 TLIOIDRANPWSACSAALITDLDGSHGDCLLDQPSKPISLPEDLPFGASYTLSQOCELAF 240
Db 193 MSLNLDHSPWSPCSAYMITSLDNGHGECLMDKPNQIQLPGDLPTGLTDANRQCQTF 252
QY 241 GVCSKPCP-YMQYCTKLWCTGAKQWCVQTRHFPWADTSGEGKCLKLGACVHERNLN 299
Db 253 GEDSKHCPDAASTCTLWCTGSGVLVQTKHFPWADTSGEGKCLKLGACVHERNLN 311
QY 300 KH---RVDGSAKWDYPGCSRTCGGVQVQARROCTNPTPANGKYCEGVRYKRSCLNLE 356
Db 312 KHFDTPFHSGWPGWPGWDCSRTCGGVQVQARROCTNPTPANGKYCEGVRYKRSCLNLE 371
QY 357 PCPSSASGKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYG 416
Db 372 DCPDN-NGKTFREOCEAHNEFSKASFGSGPAVEWIPKYAGVSPKDRCKLICQAKGIGYF 430
QY 417 YVLAPKVVDTGLCSPDSTSVQVQKICAGCDNGLSKRFPDKGVCGGDNKSKRVKVTGL 476
Db 431 FVLQPKVVDTGLCSPDSTSVQVQKICAGCDRIIDSKKFKDKGVCGGNGSTCKKISGS 490
QY 477 FTKPMHGYNFVAIPAGASSIDIRQRYKGLIGDDNYLALKNQSGKYLINGHVFVSAVER 536
Db 491 VTSKPGYHDVTIPTGATNIEVKQRNQRNNGSFLAKAADGTIYLNGDYTLSTLEQ 550
QY 537 DLVVKGSLRYSGTGTAVESLQASRPILPLETVESLVSGKMTPPRVRYSFYLPKPREDEK 596
Db 551 DIMYKGVWLRYSGSSAALERIRSFPLKEPLTIQVLVGNALRPKIYTFVKK- ---K 606
QY 597 SSHPKDPRGSPVLHNSVLSNQVEQDPRPARVAGSWGPCSASGSGLQKRAVDCRG 656
Db 607 ES-----FNAIPTFS-----AWVIEWEGCSKSCELGWQRRLVECRD 643
QY 657 SAGORTVPACDAAH--RPVETOACGE-PCPTWELSAWSPCSKSCGRGFORRLSKCVHGG 713
Db 644 INQO---PASECAKEVKAPEASTRPCADHPCPQWOLGEWSSCSKTCGKYKRLKCLSHDG 700
QY 714 RLLARQDCNLRKPKQE-LDFCVLRPC 738
Db 701 GVLSHESCDPLKPKHFIDFCTLTQC 726
```

RESULT 5

```
US-09-445-023A-1
; Sequence 1, Application US/09445023A
; Patent No. US20020119167A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiji
; APPLICANT: Hakozaiki, Michinori
; APPLICANT: Ishioke, Keiko
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
```

; TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/09/445,023A
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-445-023A-1

Query Match 55.9%; Score 2260; DB 10; Length 727;
Best Local Similarity 54.3%; Pred. No. 3.8e-157;
Matches 405; Conservative 129; Mismatches 172; Indels 40; Gaps 11;
QY 1 FVSIPTVETLVVADESVMKFGADLEHYLLTLTAARLYRHPSILNPINIVVVKVLL 60
DB 13 FVSPRYVETMLVADQSMADFGSLGKHYLLTLFVAARFYKHPKPSIRNSVSLVVKILVI 72
QY 61 RDRDSGPKVTGNAALTNRNFCAMOKLNKYSKHPYWDTAIFTRQDLCGATTCDTLGM 120
DB 73 HDEQKGPVETISNAALTNRNFCAMOKLNKYSKHPYWDTAIFTRQDLCGATTCDTLGM 132
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNPKVCEVFGKLRANHMSP 180
DB 133 ADVGTVCDPSRSCSVIEDDGLQAFTTAHELGHVFNMPHDNPKVCEVFGKLRANHMSP 192
QY 181 TLTIQIDRANPWSACSAIIITDFDSHGDCLLDQPSKPSILPDELPGASYTLSSQOCELAF 240
DB 193 MLSNLDHSQWPSPCSAYMITSFLDNGHGECIMDKPQNPQLPGDLPGLTYLDANRQCOFTF 252
QY 241 GVGSKPCP-YMQYCTKLWCTGKAGQWVCOTRHPFPWADGTSCEGKLCCLKGACVERHNLN 299
DB 253 GEDSKHCPDAASTCTLMWCTGSGVLVCQTKHFPWADGTSCEGKLCCLKGACVERHNLN 311
QY 300 KH---RVDSGMAKWDPPGCSRTCGGQVLARRQCTNPTPANGGKYCEGVRYKRSCLNE 356
DB 312 KHFTDPHGSWGPWPGWDCSRTCGGQVYTMRECONPVPKNGGKYCEGVRYKRSCLNE 371
QY 357 PCPSASGKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYGF 416
DB 372 DCPDN-NKGTFRFEOCEAHNEFSKASFGSGPAVEWIPKYAGVSPKRCCKLICQAKGIGYF 430
QY 417 YVLAPKVVVDGTLCPDSTSVCOGKCIKAGCDNGLSKRFRDCKGCGGDNKSKCKVTGL 476
DB 431 FVLQPKVVDGTPCSPDSTSVCOGQCVKAGCDRIIDSKKFKDCKGCGGNGSTCKKISGS 490
QY 477 FTKPMHGYNFVAIPAGASSIDIRQYKGLIGDDNYLALKNSQGYLLNGHVFVSAVER 536
DB 491 VTSAPKGVHDIVTPTCATNIEVKQRNQRNNGSFLAIAADGTVILNGDVTLSLEQ 550
QY 537 DLVVKGSLLRYSGTGTAVESLQASRPILPLTVESLVSGKMTPPRVRYSYLFPKPREDK 596
DB 551 DIMYKGVLYRSGSAALEIRFSPLKEPLTIQVLTGVALPKITYTFVAKK---K 606
QY 597 SSHPKDPRGPSVLHNSVLSLSNQVEQDDPPARVAGSWGPCSSASGSLQKRAVDCRG 656
DB 607 ES-----FNATPTFS-----AWVIEEMGECSKSGELGWQRRLVECRD 643
QY 657 SAGQRTVPACDAH--RPVETQAGE-PCPTWELSAWSPCKSGRFGRRSLKCVGHGG 713
DB 644 INQ---PASECAKEVPASTRPCADHPCPQWLGWSSCSKTCGKGYKRSRLKCLSHDG 700
QY 714 RLLARDQCNLHKPQE-LDFCVLRPC-738
DB 701 GVLSHESCDPLKPKHFDICTLTQC 726

RESULT 6
US-10-097-597-12

; Sequence 12, Application US/10097597
; Publication No. US20030022352A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inocuchi, Eiichi
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-097-597-12

Query Match 55.9%; Score 2259; DB 9; Length 727;
Best Local Similarity 53.6%; Pred. No. 4.5e-157;
Matches 400; Conservative 131; Mismatches 175; Indels 40; Gaps 10;
QY 1 FVSIPTVETLVVADESVMKFGADLEHYLLTLTAARLYRHPSILNPINIVVVKVLL 60
DB 13 FVSPRYVETMLVADQSMADFGSLGKHYLLTLFVAARFYKHPKPSIRNSVSLVVKILVI 72
QY 61 RDRDSGPKVTGNAALTNRNFCAMOKLNKYSKHPYWDTAIFTRQDLCGATTCDTLGM 120
DB 73 YEEQKGPVETISNAALTNRNFCAMOKLNKYSKHPYWDTAIFTRQDLCGATTCDTLGM 132
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNPKVCEVFGKLRANHMSP 180
DB 133 ADVGTVCDPSRSCSVIEDDGLQAFTTAHELGHVFNMPHDNPKVCEVFGKLRANHMSP 192
QY 181 TLTIQIDRANPWSACSAIIITDFDSHGDCLLDQPSKPSILPDELPGASYTLSSQOCELAF 240
DB 193 MLSNLDHSQWPSPCSAYMITSFLDNGHGECIMDKPQNPQLPGDLPGLTYLDANRQCOFTF 252
QY 241 GVGSKPCP-YMQYCTKLWCTGKAGQWVCOTRHPFPWADGTSCEGKLCCLKGACVERHNLN 299
DB 253 GEDSKHCPDAASTCTLMWCTGSGVLVCQTKHFPWADGTSCEGKLCCLKGACVERHNLN 311
QY 300 KH---RVDSGMAKWDPPGCSRTCGGQVLARRQCTNPTPANGGKYCEGVRYKRSCLNE 356
DB 312 KHFTDPHGSWGPWPGWDCSRTCGGQVYTMRECONPVPKNGGKYCEGVRYKRSCLNE 371
QY 357 PCPSASGKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYGF 416
DB 372 DCPDN-NKGTFRFEOCEAHNEFSKASFGNEPTVETPKYAGVSPKRCCKLTCEAKGIGYF 430
QY 417 YVLAPKVVVDGTLCPDSTSVCOGKCIKAGCDNGLSKRFRDCKGCGGDNKSKCKVTGL 476
DB 431 FVLQPKVVDGTPCSPDSTSVCOGQCVKAGCDRIIDSKKFKDCKGCGGNGSTCKKMSGI 490
QY 477 FTKPMHGYNFVAIPAGASSIDIRQYKGLIGDDNYLALKNSQGYLLNGHVFVSAVER 536
DB 491 VTSAPKGVHDIVTPTCATNIEVKQRNQRNNGSFLAIAADGTVILNGDVTLSLEQ 550
QY 537 DLVVKGSLLRYSGTGTAVESLQASRPILPLTVESLVSGKMTPPRVRYSYLFPKPREDK 596
DB 551 DIMYKGVLYRSGSAALEIRFSPLKEPLTIQVLTGVALPKITYTFVAKK---K 608
QY 597 SSHPKDPRGPSVLHNSVLSLSNQVEQDDPPARVAGSWGPCSSASGSLQKRAVDCRG 656

```
Db 609 -----FNAITPFS-----EWIEEWGECSTKTCGSGWQRVVOCRD 643
QY 657 SAGQRTVPACDAAH--RPVETQACGE--PCPTWELSAWSPSCSKSGRGRFORSLKACVGHGG 713
Db 644 INGH---PASECAKEVKPASTRPCADLPCHPWQVGDWSPSCSKTGKGYKKRTLKCVSHDG 700
QY 714 RLLARDQCNLHRKPOE-LDFCVLRPC 738
Db 701 GVLNSCSDPLKKPKHYIDFCTLTQC 726

RESULT 7
US-10-097-580-12
; Sequence 12, Application US/10097580
; Publication No. US20030032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-097-580-12
```

```
Query Match 55.9%; Score 2259; DB 9; Length 727;
Best Local Similarity 53.6%; Pred. No. 4.5e-157;
Matches 400; Conservative 131; Mismatches 175; Indels 40; Gaps 10;

QY 1 FVSPRYVETLVVADESVMKVFHGADLEHYLLTLATAARLYRHPISILNINIVVKVLLL 60
Db 13 FVSSPRVETMLVADQSMADFHGSLKHYLLTLFSAARFYKHPISIRNSISLVVKILVI 72
QY 61 RDRSGPKVTGNAALTLRNFCQAWQKLNKYSDKHPEYWDTAILFTRODLCGATCTDLGM 120
Db 73 YEEQKGPVETVNAALTLRNFCQKHNPSDRDPHEHYDTAILFTRODLCGSHCTDLGM 132
QY 121 ADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKCEVEFGKLRANHMSP 180
Db 133 ADVGTVCDDPSRSCSVIEDDGLQAAFTTAHELGHVFNMPHDHAKHCSLNGVSGDSLHMAS 192
QY 181 TLIQIDRANPWSACSAIITDLSGHDCLLDQPSKPISLPEDLPQASVYTLSSQCEELAF 240
Db 193 MLSSLDHSDQWSPSCSAVMYVTSFLDNGHCEGLMDKPNQPIKLPDLPGLTYLDANRQCQFTF 252
QY 241 GVSKPCP-VMQYCTKLWCTGKAGQMVQCTRHFPWADGTSCGEGKLCACVERHNLN 299
Db 253 GEERKCPDAASTCTTLWCTGTSGLLVCQTKHFPWADGTSCGEGKWCVSGKVCNKTDM- 311
QY 300 KH---RVDGSMKWDPYGPCSRTCGGGVQLARROCTNPTPANGKCYCEGVRYKVRSONLE 356
Db 312 KHFAFTPVHSGWGPWGPWGDSCRTCGGGVQVYTMRECDNPVPNGKCYCEGARVIRSCNIE 371
QY 357 PCPSASGKSFREQCEAFNGYNHSTNRLTLAVAWPKYSGVSPDRCKLICRANGTGYF 416
Db 372 DCPDN-NGKTFREQCEAHNEFSKASFGNEPTVEWTFKYAGVSPKDRCKLTCEAKGIGYF 430
```

```
QY 417 YVLAPKVVDTGLTSPDSTSVCGQKICAGCDGNLGSKKRFRDKCGVCGGDNKSKKVTGL 476
Db 431 FVLQPKVVDGTPCSPDSTSVCGQCVQKACDRIIDSKKFKDCGCGVCGGNGSTCKKMSGI 490
QY 477 FTKPMHGYNVVAIPAGASSIDIRQGYKGLIGDDNYLALKNQSGKYLLNGHFVWSAVER 536
Db 491 VTSTRFGYHDIVTPAGATNIEVKHRNGSRNNGSFLAIRAADGTIYILNGNFTLTLEQ 550
QY 537 DLVVGSLRLRYSGTGTAVESIOASRPILTEPLTVEVLSVSKMTTPRVRYSFYLPKEPREDK 596
Db 551 DLTYSKGLVLRYSGSAALERISFSLKEPLTIQVLMVGHALRPKIKFTFYFMKKKTES-- 608
QY 597 SSHPKDPRGSPSVLHNSVLSLSNQVEQPDPRPARWAGVSGWGPCSACSGSLQKRAVDCRG 656
Db 609 -----FNAITPFS-----EWIEEWGECSTKTCGSGWQRVVOCRD 643
QY 657 SAGQRTVPACDAAH--RPVETQACGE--PCPTWELSAWSPSCSKSGRGRFORSLKCVGHGG 713
Db 644 INGH---PASECAKEVKPASTRPCADLPCHPWQVGDWSPSCSKTGKGYKKRTLKCVSHDG 700
QY 714 RLLARDQCNLHRKPOE-LDFCVLRPC 738
Db 701 GVLNSCSDPLKKPKHYIDFCTLTQC 726

RESULT 8
US-09-445-023A-12
; Sequence 12, Application US/09445023A
; Patent No. US20020119167A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/09/445,023A
; CURRENT FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-09-445-023A-12

Query Match 55.9%; Score 2259; DB 10; Length 727;
Best Local Similarity 53.6%; Pred. No. 4.5e-157;
Matches 400; Conservative 131; Mismatches 175; Indels 40; Gaps 10;

QY 1 FVSPRYVETLVVADESVMKVFHGADLEHYLLTLATAARLYRHPISILNINIVVKVLLL 60
Db 13 FVSSPRVETMLVADQSMADFHGSLKHYLLTLFSAARFYKHPISIRNSISLVVKILVI 72
QY 61 RDRSGPKVTGNAALTLRNFCQAWQKLNKYSDKHPEYWDTAILFTRODLCGATCTDLGM 120
Db 73 YEEQKGPVETVNAALTLRNFCQKHNPSDRDPHEHYDTAILFTRODLCGSHCTDLGM 132
QY 121 ADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKCEVEFGKLRANHMSP 180
Db 133 ADVGTVCDDPSRSCSVIEDDGLQAAFTTAHELGHVFNMPHDHAKHCSLNGVSGDSLHMAS 192
QY 181 TLIQIDRANPWSACSAIITDLSGHDCLLDQPSKPISLPEDLPQASVYTLSSQCEELAF 240
Db 193 MLSSLDHSDQWSPSCSAVMYVTSFLDNGHCEGLMDKPNQPIKLPDLPGLTYLDANRQCQFTF 252
QY 241 GVSKPCP-VMQYCTKLWCTGKAGQMVQCTRHFPWADGTSCGEGKLCACVERHNLN 299
```

```

Db 253 GEESKCPDAASTCTTLCWGTGSGLLVCQTKHFPWADGTSCGEGKWCVCNKTMDM- 311
QY 300 KH---RVDSWAKWDPYPCSTCGGQVQLARRCTNTPANGKYCGEVRYKRSCLNLE 356
Db 312 KHFAIPVHSGWGPWPGWDCSRTCGGQVQYTMRECDNPVPKNGKYCGEVRYKRSCLN 371
QY 357 PCPSASGKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYGF 416
Db 372 DCPDN-NGKTFREOCEAHNEFSKASFGNEPTVETPKYAGVSPKDRCKLTCEAKGIGYF 430
QY 417 YVLAPKVVDTGTLCPDSTSVQVQKICRAGDGNLGSKKRDKCGVCGDGNKSKKVTGL 476
Db 431 FVLQPKVVDGTGTLCPDSTSVQVQKICRAGDGNLGSKKRDKCGVCGDGNKSKKVTGL 490
QY 477 FTKPMHGYNFVVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQKYLKNGHVFVSAVER 536
Db 491 VTSTRPGYHDIVTPAGATNIEVKHNRQGRNNGSFLAIRADGTYLILNGFTLSTLEQ 550
QY 537 DLVKGSLRYSGTGTAVESLQASRPILPLTVEVLSVGKMTPPRRVRYSFYLPKPEPREDK 596
Db 551 DLTKYGTVLYRSGSSAALRISFSLKEPLTIQVLMVGHALRPKIKFTYFMKKKTES-- 608
QY 597 SSHPKDPRGSPVLSNLSNOVEQDDPPARVWAGSWGPCSCSGGLQKRAVDCRG 656
Db 609 -----FNAIPTFS-----EWVIEEWGECSTCGSGWQRRVQCRD 643
QY 657 SAGQRTVPACDAAH--RPVETOACGE-PCPTWELSAWSPCSKSGRFFORSLKCVGHGG 713
Db 644 INGH---PASECAKEVKPASTRPCADLPCPHWQVGDWSPCSKTCGKYKRTLCVSHDG 700
QY 714 RLLARDQCNLHRKPQE-LDFCVLRPC 738
Db 701 GVLNESCDDLKPKHYIDFCTLTQC 726

RESULT 9
US-09-321-987B-4
; Sequence 4, Application US/09321987B
; Patent No. US20020102210A1
; GENERAL INFORMATION:
; APPLICANT: Kimble, Judith E
; APPLICANT: Billeloch, Robert H
; TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
; FILE REFERENCE: 960296.95386
; CURRENT APPLICATION NUMBER: US/09/321,987B
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087,170
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/129,023
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Murine
US-09-321-987B-4

Query Match 55.8%; Score 2257; DB 10; Length 950;
Best Local Similarity 53.6%; Pred. No. 8.8e-157;
Matches 400; Conservative 130; Mismatches 176; Indels 40; Gaps 10;

QY 1 FVSIPRYVETLVVADESNVFKFHGADLEHYLLTLATAARLYRHPSILNPINIVVVKVLL 60
Db 237 FVSSPRVETMLVADQSMADFHGSLKHYLLTLFSAARFYKHPKSRISLSLVVVKILVI 296
QY 61 RDRSGPKVTGNAALTNRNCAWQKLNKYSKHPYWDTAITLFTRODLGATTCDTGLM 120
Db 237 YEEQKGPVETNAALTNRNCAWQKLNKYSKHPYWDTAITLFTRODLGATTCDTGLM 356
QY 121 ADVGTWCDPKRSCSVIEDDGLPFAFTTAHELGHVFNPHDNPVKNVYCEEVFGKLRANHMSP 180

```

```

Db 357 ADVGTWCDPKRSCSVIEDDGLPFAFTTAHELGHVFNPHDNPVKNVYCEEVFGKLRANHMSP 416
QY 181 TLQIDRANPWSACSAIITDFDSDGHGDCLLDPQSPKPSLDPEDLPASVTLSCQCELAFL 240
Db 417 MLSLSDHSQSPSCSAIITDFDSDGHGDCLLDPQSPKPSLDPEDLPASVTLSCQCELAFL 476
QY 241 GVGSKPCP-YMQYTKLWCTGKAKGQWVCOTRHPFPWADGTSCGEGKLCGLKAGACVERHNLN 299
Db 477 GEESKCPDAASTCTTLCWGTGSGLLVCQTKHFPWADGTSCGEGKWCVCNKTMDM- 535
QY 300 KH---RVDSWAKWDPYPCSTCGGQVQLARRCTNTPANGKYCGEVRYKRSCLNLE 356
Db 536 KHFAIPVHSGWGPWPGWDCSRTCGGQVQYTMRECDNPVPKNGKYCGEVRYKRSCLN 595
QY 357 PCPSASGKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRANGTYGF 416
Db 596 DCPDN-NGKTFREOCEAHNEFSKASFGNEPTVETPKYAGVSPKDRCKLTCEAKGIGYF 654
QY 417 YVLAPKVVDTGTLCPDSTSVQVQKICRAGDGNLGSKKRDKCGVCGDGNKSKKVTGL 476
Db 655 FVLQPKVVDGTGTLCPDSTSVQVQKICRAGDGNLGSKKRDKCGVCGDGNKSKKVTGL 714
QY 477 FTKPMHGYNFVVAIPAGASSIDIRQRYKGLIGDDNYLALKNSQKYLKNGHVFVSAVER 536
Db 715 VTSTRPGYHDIVTPAGATNIEVKHNRQGRNNGSFLAIRADGTYLILNGFTLSTLEQ 774
QY 537 DLVKGSLRYSGTGTAVESLQASRPILPLTVEVLSVGKMTPPRRVRYSFYLPKPEPREDK 596
Db 775 DLTKYGTVLYRSGSSAALRISFSLKEPLTIQVLMVGHALRPKIKFTYFMKKKTES-- 832
QY 597 SSHPKDPRGSPVLSNLSNOVEQDDPPARVWAGSWGPCSCSGGLQKRAVDCRG 656
Db 833 -----FNAIPTFS-----EWVIEEWGECSTCGSGWQRRVQCRD 867
QY 657 SAGQRTVPACDAAH--RPVETOACGE-PCPTWELSAWSPCSKSGRFFORSLKCVGHGG 713
Db 868 INGH---PASECAKEVKPASTRPCADLPCPHWQVGDWSPCSKTCGKYKRTLCVSHDG 924
QY 714 RLLARDQCNLHRKPQE-LDFCVLRPC 738
Db 925 GVLNESCDDLKPKHYIDFCTLTQC 950

RESULT 10
US-10-163-316-7
; Sequence 7, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MP101-025P1RNM
; CURRENT APPLICATION NUMBER: US/10/163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-163-316-7

Query Match 55.8%; Score 2257; DB 9; Length 968;
Best Local Similarity 53.6%; Pred. No. 9e-157;
Matches 400; Conservative 130; Mismatches 176; Indels 40; Gaps 10;

QY 1 FVSIPRYVETLVVADESNVFKFHGADLEHYLLTLATAARLYRHPSILNPINIVVVKVLL 60
Db 254 FVSSPRVETMLVADQSMADFHGSLKHYLLTLFSAARFYKHPKSRISLSLVVVKILVI 313
QY 61 RDRSGPKVTGNAALTNRNCAWQKLNKYSKHPYWDTAITLFTRODLGATTCDTGLM 120

```



```
Db 314 YEQKGPVTSNAALTRNFCAWKQNSPSDRDPEHYDTAILFTRODLCGSHFTCDTLGM 373
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNDVYKCEEVFGKLRANHMSP 180
Db 374 ADVGTVCDPKSCSVIEDDGLQAFTTAHELGHVFNPHDNDVYKCEEVFGKLRANHMSP 433
QY 181 TLQIDRANPWSACSAIITDFDLSHGDCDLDQSPKIPSLPDLPGASYTLSQOCELAF 240
Db 434 MSLSDHSPQWPCSAVMYTSFLDNGHGBCLMDKPNQIKLPDLPGTLYDANRQCQFTF 493
QY 241 GVGSKPCP-YMOYCTKLWCTGKAGQWVCQTRHFPWADGTSCGEGKLCGLKACVERHNLN 299
Db 494 GEESKPCDAASTCTLWCTGSGLLVCQTKHFPWADGTSCGEGKLCGLKACVERHNLN 552
QY 300 KH---RVDSGNAKWDYPGCSRTCGGVQVLARRQCTNPTPANGKYCEGVRYKRSCLNE 356
Db 553 KHFATPVHSGWPGWPGWDCSRTCGGVQVYTMRECDNPVPKNGKYCEGVRYKRSCLNE 612
QY 357 PCPSASGKSFREOCEAFNGYNHSTNRLTLAVAWVPKYSGVSPRDKKLCICRANGTYGF 416
Db 613 DCPDN-NGKTFREOCEAHNEFSKASFGNEPTVEWTPKYAGVSPKDRCKLTCEAKGIYF 671
QY 417 YVLAPKVVDTGTLSPDSTSVQGVQKICAGQDGNLGSKKRFDKCGVCGDNKCKKVTGL 476
Db 672 FVLQPKVVDTGTLSPDSTSVQGVQKICAGQDRIIDSKKFDKCGVCGDNKCKKVTGL 731
QY 477 FTKPMHGVNFVAIPAGASSIDIRQYKGLIGDDNYLALKNSQGYLLNGHFVVSAYER 536
Db 732 VTSTRPGVHDITVTPAGATNIEVKQRNQRSGRNSFLAIRAADGTYLNGTFTLSLEQ 791
QY 537 DLVVKGLSLRYSGTGAVESLOASRPILPLVEVLVSVGKMTPPRVVRSFYLPKPREDK 596
Db 792 DLTGKGVILRYSGSSAALERISFSPKLEPTIQVLTGNALRPKIKITYFVYKK- 846
QY 597 SSHPKDPRGPSVYLHNSVLSNQVEQDPPRPPARWAGSWGSPCSASCGSLQKRAVDCRG 656
Db 850 -----FNAIPTFS-----AWVIEEWGECSSKCELGWQRRLVECDR 883
QY 657 SAGQRTVPACDAAH--RPVETOACGE-PCPTWELSAWSPCSKSGRGQRRLKCVGHGG 713
Db 885 INGH---PASECAKEVPKASTRPCADLPCPHWQVGDWSPCSKSGKGYKRLKCLSHDG 941
QY 714 RLLARDQCNLHRKPOE-LDFCVLRPC 738
Db 942 GVLNESCDELKPKHIDFCTLTQC 967
```

```
RESULT 11
US-10-105-929-2
; Sequence 2, Application US/10105929
; Patent No. US20020137142A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/10/105,929
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/130,491
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-105-929-2
```

```
Query Match 55.8%; Score 2254; DB 12; Length 967;
Best Local Similarity 54.0%; Pred. No. 1.5e-156;
Matches 403; Conservative 131; Mismatches 172; Indels 40; Gaps 11;

QY 1 FVSIPIRYVETLVVADSWKFGHADLEHYLLTLLATAARLYRHPSTILNPINIVVKKVLL 60
Db 253 FVSHRYVETLVVADSWKFGHADLEHYLLTLLATAARLYRHPSTILNPINIVVKKVLL 312
QY 61 RDRSDGPKVTGNAALTRNFCAWKQNSPSDRDPEHYDTAILFTRODLCGSHFTCDTLGM 120
Db 313 HDQKQGEVTSNAALTRNFCAWKQNSPSDRDPEHYDTAILFTRODLCGSHFTCDTLGM 372
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNDVYKCEEVFGKLRANHMSP 180
Db 373 ADVGTVCDPKSCSVIEDDGLQAFTTAHELGHVFNPHDNDVYKCEEVFGKLRANHMSP 432
QY 181 TLQIDRANPWSACSAIITDFDLSHGDCDLDQSPKIPSLPDLPGASYTLSQOCELAF 240
Db 433 MSLSDHSPQWPCSAVMYTSFLDNGHGBCLMDKPNQIKLPDLPGTLYDANRQCQFTF 492
QY 241 GVGSKPCP-YMOYCTKLWCTGKAGQWVCQTRHFPWADGTSCGEGKLCGLKACVERHNLN 299
Db 493 GEESKPCDAASTCTLWCTGSGLLVCQTKHFPWADGTSCGEGKLCGLKACVERHNLN 551
QY 300 KH---RVDSGNAKWDYPGCSRTCGGVQVLARRQCTNPTPANGKYCEGVRYKRSCLNE 356
Db 552 KHFATPVHSGWPGWPGWDCSRTCGGVQVYTMRECDNPVPKNGKYCEGVRYKRSCLNE 611
QY 357 PCPSASGKSFREOCEAFNGYNHSTNRLTLAVAWVPKYSGVSPRDKKLCICRANGTYGF 416
Db 612 DCPDN-NGKTFREOCEAHNEFSKASFGNEPTVEWTPKYAGVSPKDRCKLTCEAKGIYF 670
QY 417 YVLAPKVVDTGTLSPDSTSVQGVQKICAGQDGNLGSKKRFDKCGVCGDNKCKKVTGL 476
Db 671 FVLQPKVVDTGTLSPDSTSVQGVQKICAGQDRIIDSKKFDKCGVCGDNKCKKVTGL 730
QY 477 FTKPMHGVNFVAIPAGASSIDIRQYKGLIGDDNYLALKNSQGYLLNGHFVVSAYER 536
Db 731 VTSKPEYHDIITPTGATNIEVKQRNQRSGRNSFLAIRAADGTYLNGTFTLSLEQ 790
QY 537 DLVVKGLSLRYSGTGAVESLOASRPILPLVEVLVSVGKMTPPRVVRSFYLPKPREDK 596
Db 791 DLTGKGVILRYSGSSAALERISFSPKLEPTIQVLTGNALRPKIKITYFVYKK- 846
QY 597 SSHPKDPRGPSVYLHNSVLSNQVEQDPPRPPARWAGSWGSPCSASCGSLQKRAVDCRG 656
Db 847 ES-----FNAIPTFS-----AWVIEEWGECSSKCELGWQRRLVECDR 883
QY 657 SAGQRTVPACDAAH--RPVETOACGE-PCPTWELSAWSPCSKSGRGQRRLKCVGHGG 713
Db 884 INGH---PASECAKEVPKASTRPCADLPCPHWQVGDWSPCSKSGKGYKRLKCLSHDG 940
QY 714 RLLARDQCNLHRKPOE-LDFCVLRPC 738
Db 941 GVLNESCDELKPKHIDFCTLTQC 966
```

```
RESULT 12
US-09-918-171A-9
; Sequence 9, Application US/09918171A
; Patent No. US20020110894A1
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirshata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/04193
; CURRENT APPLICATION NUMBER: US/09/918,171A
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 09/369,364
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 9
; LENGTH: 905
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-8
US-09-918-171A-9

Query Match          47.3%; Score 1911.5; DB 10; Length 905;
Best Local Similarity 50.6%; Pred. No. 1.5e-131;
Matches 352; Conservative 127; Mismatches 185; Indels 31; Gaps 13;

QY 1 FVSIPRYVELTVVADSMVYKFGADLHYLLTLLATARLYRHPSTLNPINIVVVKVLL 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 FVSEARFETLLVADASMAAFYGTDLQNHILTVMSMAARIYKHPSTRNSVNLVVVKLV 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 RDRDSGPKVTGNALTLNFCAMOKLNVSKDHPYWDATILFTRODLGCG-ATTCDTLG 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 EKERWGPVSDNGGLTLNFCNSWRFRNPKPSDRHPHYDTAILFTQNFCGGEQCDTLG 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 120 MADVGTMCOPKPRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEEVFGKLRANHMS 179
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 MADVGTICDPKSCSVIKDEGLQAAYTLAHELGHVLSMHPDDSKPCVRLFGPMGVYHMA 408

QY 180 PTLIQIDRANPWACSAAIITDFLDGSHGDCILLDQPSKIPISLPEDLPAS--YTLSQOCE 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 PFEIHNKTLWPSPCPSAVYLTLLDGHGDCILLDAPTSVLPFTGLPGHSTLYELDQOQCK 468

QY 238 LAFGVGSKPCP---YMQYCTKLWCTKAKGQWVCOTRH--FWADGTSCGEGKLCCLKGAC 292
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 QIFGPDFRHCNTSVEDICVQLCARHDSDEPICHTKNSLLWADTGPCGPHCLDGGSC 528

QY 293 VERHNLN--KHRYDGSWAKWDYGPCSRTCGGQVQLARRQCTNPTPANGKCYCEGVYRVKY 350
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 VLKEDVENKAVVDGWPWRPWGQCSRTCGGQIQPSNRECDNPMQNGRFGCLGERVKY 588

QY 351 RSNLEPCPSASGSKFRECEAFNGYHNHRTNLTLAVWPVKYSGVSPRDKCKLICRA 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 589 QSCNTEPC--PNKGKFRQOCEKYNAYNH-TDLGNFLQWVPKYSGVSPRDKCKLICRA 645

QY 411 NGTYFVVLAPKVVVDGTLCPSPDSTSVCGVKCIKAGCDNGLSKKRFDCGCGGDNKSC 470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 646 RGRSEKVFPEAKVIDGTLCGPDFTLSCVRQCVKAGCDHVNSPKLDRKCGVGGKGATC 705

QY 471 KKVTLGFTKPMHGYNFVVAIPAGASSIDIRQRYKGLIGDDNVLAKNSGKYLNGHFV 530
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 706 RKISGFTPEYGYNDIVTIPAGATNIDVKQRSHPGVRNDGYSYALKTANGQYLLNGNLA 765

QY 531 VSAVERDLVKGSLRYSGTGTAVESLQASRPILPTVEVLVS-GKMTPPRVRYSFYLP 589
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 ISATEQDILVKGTLTKYSGSMATLERLQSFQALPELTVQLLVSGEVFPKRYTFYFP 825

QY 590 KEPREDKSHPKDPRGPSVLHNSVLSNQVEQPDPRPARWAGSWGPCASCSGLQK 649
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 826 ND--MDFSVQSKERATY---NIIQSL-----PSAEWVLGDWSECPSTCRGSQWR 870

QY 650 RAVDCRGSAGQRTVPACDAARHPVETQACG-EPCC 683
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 871 RTVECDRPSGQAS-DTCDEALKPEDAKPCGSPCP 904

RESULT 13
US-09-803-589-2
; Sequence 2, Application us/09803589
; Patent No. US20020112251A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
; TITLE OF INVENTION: USES
; FILE REFERENCE: 07334-325001
; CURRENT APPLICATION NUMBER: US/09/803,589
; CURRENT FILING DATE: 2001-03-09
```

```
; PRIOR APPLICATION NUMBER: US 09/128,709
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 60/054,645
; PRIOR FILING DATE: 1997-08-04
; PRIOR APPLICATION NUMBER: US 09/130,491
; PRIOR FILING DATE: 1998-08-06
; PRIOR APPLICATION NUMBER: US 60/054,966
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 09/388,280
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 09/388,279
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-803-589-2

Query Match          47.0%; Score 1901; DB 10; Length 608;
Best Local Similarity 49.3%; Pred. No. 5.3e-131;
Matches 339; Conservative 114; Mismatches 146; Indels 88; Gaps 8;

QY 57 VLLLRDRDSGPKVTGNALTLNFCAMOKLNVSKDHPYWDATILFTRODLGCGATTC 116
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 ILVIHDEQKPEVTSNAALTNRNFCNWKQHNPPSDRAEHYDTAILFTRODLGSGQTC 63

QY 117 TLGMADVTCMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEVFCCKLRANH 176
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 TLGMADVTCVCDPKRSCSVIEDDGLQAFAFTTAHELGHVFNPHDDDAKQCASLNGVNDSH 123

QY 177 MMSPTLIQIDRANPWACSAAIITDFLDGSHGDCILLDQPSKIPISLPEDLPASGYTL 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 MWASMLNLDHISQWSPCSAYWITSFLDNGHGECLMDKDPONPIQLPGDLPTSYDANRQC 183

QY 237 ELAFGVGSKPCP-YMQYCTKLWCTKAKGQWVCOTRHFPWADGTSCGEGKLCCLKGACVER 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 QFTGEDSKHCPDAASTCTSLWCTGTSVCGVLCQTKHFPWADGTSCGEGKWCINGKCVNK 243

QY 296 HNLNKH---RVDGSAKWDPYGPCSRTCGGQVQLARRQCTNPTPANGKCYCEGVYRVKYS 352
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 244 TD-RKHEFTPPHSGWMGMPWGDCTRTCGGVQYTMRECDNPPKNGKCYCEGKRVYRS 302

QY 353 CNLEPCPSASGSKSPREEQCEAFNGYHNHRTNLTLAVAWPKYSGVSPRDKCKLICRANG 412
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 CNLEDCPDN-NGKTFREEQCEAHNEFSKASFGSGGPAVEWIPKYAGVSPKDRCKLICQAKG 361

QY 413 TGTYFVVLAPKVVVDGTLCPSPDSTSVCGVKCIKAGCDNGLSKKRFDCGCGGDNKSCCK 472
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 IGYFFVLOPKVVVDGTPCSPDSTSVCGVQCCKAGCDRIIDSKKRFDCGCGGNGSTCKK 421

QY 473 VTGLFTKPMHGYNFVVAIPAGASSIDIRQRYKGLIGDDNVLAKNSGKYLNGHFVYS 532
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 ISGVSISAKPGYHDIITPTGATNIEVKQRNGRNSGFLAIKADGYTLNGDYTLS 481

QY 533 AVERDLVKGSLRYSGTGTAVESLQASRPILPTVEVLVSVGKMTPPRVRYSFYLPKEP 592
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 TLEQDIMYKGVVLYRSGSAAERIRSFSPLEPTIQVLTGVGNALRPKIRYTFYEVKKK- 540

QY 593 REDKSSHPKDPGRGPSVLHNSVLSNQVEQPDPRPARWAGSWGPCASCSGGLQKRAV 652
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 ---KES-----FNAITFS----- 551

QY 653 DCRGSAGQRTVPACDAARHPVETQACGEPCTWELSAWSPCSKSGRGFORSLKCVGHG 712
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 552 -----AWIIEGECCKTCGKGYKRSKLSKLSHD 580

QY 713 GRLLARDQCNLHRKPOE-LDFCVLRPC 738
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 581 GGVLSHESCDPLKPKKHFIDFCTMAEC 607
```

RESULT 14
US-09-803-589-8
; Sequence 8, Application US/09803589
; Patent No. US20020112251A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
; TITLE OF INVENTION: PROGNOSTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC AND OTHER
; TITLE OF INVENTION: USES
; FILE REFERENCE: 07334-325001
; CURRENT APPLICATION NUMBER: US/09/803,589
; CURRENT FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: US 09/128,709
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 60/054,645
; PRIOR FILING DATE: 1997-08-04
; PRIOR APPLICATION NUMBER: US 09/130,491
; PRIOR FILING DATE: 1998-08-06
; PRIOR APPLICATION NUMBER: US 60/054,966
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 09/388,280
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: US 09/388,279
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-803-589-8

Query Match 47.0%; Score 1900; DB 10; Length 608;
Best Local Similarity 49.3%; Pred. No. 6.3e-131;
Matches 339; Conservative 114; Mismatches 146; Indels 88; Gaps 8;

QY 57 VLLLRDRSGPKVTGNAALTLRNFCAWQKLNKVSXKHPEYWDTAITLFTRODLGATTC 116
Db 4 ILVHDEQKGPVTSNAALTLRNFCAWQKLNKVSXKHPEYWDTAITLFTRODLGATTC 63
QY 117 TLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNNVYCEVFGKLRANH 176
Db 64 TLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNNVYCEVFGKLRANH 123
QY 177 MMSPTLIQIDRANPWSACSAAITDPLDSHGDCILDDPSKPISLPDLPGASYTL 236
Db 124 MMSPTLIQIDRANPWSACSAAITDPLDSHGDCILDDPSKPISLPDLPGASYTL 183
QY 237 ELAFGVGSKPCP-YMOYCTKLMCTGKAGQWVCQTRHFPWADGTCGEGKLCIKGACVER 295
Db 184 QFTFGEDSKHCPDAASTCTLWCTGSGVLVCQTKHFPWADGTCGEGKLCIKGACVER 243
QY 296 HNLNKH---RVDSWAKWDYPCSRCTCGGVQVQLARRQCTNPTPANGKYCEGVYKYS 352
Db 244 TD-RKHFDTPFHGSGWGMWPGDCSRTCGGVQVYTMRECDNPVKNKKYCEGVYKYS 302
QY 353 CNLEPCSSASGSKFREEOCEAFNGYNHSTNRLTLAVAVPKYSGVSPDKCLICRANG 412
Db 303 CNLEPCSSASGSKFREEOCEAFNGYNHSTNRLTLAVAVPKYSGVSPDKCLICRANG 361
QY 413 TGYFYVLPKVVVDGTLCPDSTSVCVQKCIKAGDGNLGSKKRFDKCGVCGGDNKSKK 472
Db 362 IGYFFVLQPKVVDGTLCPDSTSVCVQKCIKAGDGNLGSKKRFDKCGVCGGDNKSKK 421
QY 473 VTGLFTKPMHGVNFVAIPAGASSIDIRQYKGLIGDDNYLALNKSQKYLINGHFVYS 532
Db 422 IGSVTSAPKPGYHDIITIPIGATNIEVKQRNQRNNGSFLAIKAADGTIILNGDYTLS 481

QY 533 AVERDLVVGKSLRYSGTGTAVESLQASRPILPTEVLVSXGKMTPPRVRYSYFLPKPE 592
Db 482 TLEODIMYKGVLRYSGSSAALIRSFSPKEPLTIQVLTGNALRPLKITYFYFKKK- 540
QY 593 REDKSSHPKDPGRGVSVLHNSVLSNQVEQDDPPARWVAGSWGSPCSASCGLQKRAV 652
Db 541 ---KES-----FNAITFS----- 551
QY 653 DCRGSAGQRTVPACDAAHRPVETQACGEPCTWELSAMSPCSKSGRGFORRLKCVGHG 712
Db 552 -----AWVIEWGECSKTCGKYKRSKLCLSHD 580
QY 713 GRLLARDQCNLHRKQPB-LDFCVLRPC 738
Db 581 GGVLSHESCDPLKPKHFIDFTWAE 607
RESULT 15
US-10-105-929-13
; Sequence 13, Application US/10105929
; Patent No. US20020137142A1
; GENERAL INFORMATION:
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/10/105,929
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/130,491
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-105-929-13

Query Match 47.0%; Score 1900; DB 12; Length 608;
Best Local Similarity 49.3%; Pred. No. 6.3e-131;
Matches 339; Conservative 114; Mismatches 146; Indels 88; Gaps 8;

QY 57 VLLLRDRSGPKVTGNAALTLRNFCAWQKLNKVSXKHPEYWDTAITLFTRODLGATTC 116
Db 4 ILVHDEQKGPVTSNAALTLRNFCAWQKLNKVSXKHPEYWDTAITLFTRODLGATTC 63
QY 117 TLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNNVYCEVFGKLRANH 176
Db 64 TLGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNNVYCEVFGKLRANH 123
QY 177 MMSPTLIQIDRANPWSACSAAITDPLDSHGDCILDDPSKPISLPDLPGASYTL 236
Db 124 MMSPTLIQIDRANPWSACSAAITDPLDSHGDCILDDPSKPISLPDLPGASYTL 183
QY 237 ELAFGVGSKPCP-YMOYCTKLMCTGKAGQWVCQTRHFPWADGTCGEGKLCIKGACVER 295
Db 184 QFTFGEDSKHCPDAASTCTLWCTGSGVLVCQTKHFPWADGTCGEGKLCIKGACVER 243
QY 296 HNLNKH---RVDSWAKWDYPCSRCTCGGVQVQLARRQCTNPTPANGKYCEGVYKYS 352
Db 244 TD-RKHFDTPFHGSGWGMWPGDCSRTCGGVQVYTMRECDNPVKNKKYCEGVYKYS 302
QY 353 CNLEPCSSASGSKFREEOCEAFNGYNHSTNRLTLAVAVPKYSGVSPDKCLICRANG 412
Db 303 CNLEPCSSASGSKFREEOCEAFNGYNHSTNRLTLAVAVPKYSGVSPDKCLICRANG 361
QY 413 TGYFYVLPKVVVDGTLCPDSTSVCVQKCIKAGDGNLGSKKRFDKCGVCGGDNKSKK 472

```
Db 362 IGYFFVLQPKVVDGTPCSPDSTSVCVQGCVKAGCDRIIDSKKKFKDKGVCGGNGSTCKK 421
QY 473 VTGLTKPMHGYNEVVAIPAGASSIDIRQYKGLICDDNYLALKNSQKYLINGHFVVS 532
Db 422 ISGSVTSAPGYHDIITIPIGATNIEVKQNRQSRNNGSFLAIIKAADGTIILNGDYTLS 481
QY 533 AVERDLVVKGSLRLRYSGTGTAVESLQASRPILEPLTVEVLSVGKMTPPRVRYSFYLPKEP 592
Db 482 TLEQDIMYKGVWLRYSGSSAALERIRSFPLKEPLTIQVLTGVGNALRPKIKYTYFVKKK- 540
QY 593 REDKSSHPKDPGRGPSVLHNSVLSLNOVEQPDPRPARVAGSWGSPCSASCGLQKRAV 652
Db 541 ---KES-----FNAIPTFS----- 551
QY 653 DCRGSAGQRTVPACDAAHREVETQACGEPCTWELSAWSPCKSCGRGFQRRSLKCVGHG 712
Db 552 -----AWVIEWGECSKTCGKGYKKRSLKCLSHD 580
QY 713 GRLLARDQCNLHRKPQE-LDFCVLRPC 738
Db 581 GGVLSHESCDPLKKPKHFIDFCTMAEC 607
```

Search completed: April 29, 2003, 17:22:03
Job time : 21.4049 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 17.07 Seconds
(without alignments)
4156.253 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_950

Perfect score: 4043
Sequence: 1 FVSIPIRYVETLVVADESMVK.....DQNLHRKPQELDFCVLRPC 738

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2257	55.8	951	2 T00017	gene ADAMTS-1 prot
2	1761	43.6	837	2 T00355	hypothetical prote
3	1604	39.7	550	2 T47158	hypothetical prote
4	1567	38.8	2165	2 T21371	hypothetical prote
5	1045.5	25.9	1205	2 T18517	procollagen N-endo
6	681	16.8	1558	2 C89114	protein C37c3.6a [
7	691	16.8	2167	2 T34395	hypothetical prote
8	510.5	12.6	860	2 T16892	hypothetical prote
9	504	12.5	1444	2 T18856	angiogenesis inhib
10	493	12.2	951	2 T00260	hypothetical prote
11	412	10.2	957	2 T15976	hypothetical prote
12	411.5	10.2	1059	2 T22545	hypothetical prote
13	258.5	6.4	571	2 S24789	Jarrahagin C. precu
14	255.5	6.3	1584	2 T00026	brain-specific ang
15	253.5	6.3	1074	2 JC5928	semaphorin F precu
16	248.5	6.1	1170	1 TSHUP1	thrombospondin 1 p
17	245.5	6.1	1170	2 A40558	thrombospondin 1 p
18	244.5	6.0	419	2 A59414	metalloproteinase
19	243	6.0	609	2 S55270	catrocollastatin p
20	234.5	5.8	508	2 T22836	hypothetical prote
21	234	5.8	419	2 S41607	atrolysin A (EC 3.
22	234	5.8	952	2 T18900	disintegrin and me
23	233.5	5.8	903	2 S60257	metrin alpha - mo
24	229	5.7	549	2 S48169	metalloproteinase
25	224	5.5	617	2 S48160	metalloproteinase
26	220	5.4	416	2 A37877	hemorrhagic protei
27	219	5.4	814	2 G02390	disintegrin-like m
28	217	5.4	826	2 A60385	monocyte surface a
29	213	5.3	898	2 T14764	hypothetical prote

30 211.5 5.2 600 2 T49281
31 208.5 5.2 1172 2 A42587
32 205.5 5.1 429 2 A42972
33 201.5 5.0 407 2 S66260
34 200.5 5.0 788 2 T25061
35 199 4.9 1172 1 TSHUP2
36 198 4.9 616 2 A55796
37 197.5 4.9 411 1 HYSNFA
38 197 4.9 480 1 A30065
39 196.5 4.9 1239 2 T13809
40 195.5 4.8 203 2 A59421
41 195.5 4.8 437 2 S05478
42 195.5 4.8 660 2 S71943
43 194 4.8 1178 1 A39804
44 193.5 4.8 481 2 JC4342
45 192 4.7 789 2 S28259

fertilin alpha pre
thrombospondin 2 p
coagulation factor
metalloproteinase
hypothetical prote
thrombospondin 2 p
ecarin precursor -
triglycane (EC 3.4.
fibrinolase (EC 3.4.
probable disintegr
metalloproteinase
propeptidase - mouse
metalloproteinase
thrombospondin pre
fibrinolytic prote
androgen-regulated

ALIGNMENTS

RESULT 1

T00017

gene ADAMTS-1 protein - mouse

C.Species: Mus musculus (house mouse)

C.Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000

C.Accession: T00017

R.Kuno, K.; Lizasa, H.; Ohno, S.; Matsushima, K.

Genomics 46, 466-471, 1997

A.Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 g

A.Reference number: Z14055; MUID:98110583; PMID:9441751

A.Accession: T00017

A.Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-951 <KUN>

A.Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BAA24501.1; PID:g2809057

A.Experimental source: strain 129SVJ

C.Genetics:

A:Gene: ADAMTS-1

A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 602/1; 660/3; 719/2

C:Superfamily: thrombospondin type 1 repeat homology

F:542-598/Domain: thrombospondin type 1 repeat homology <THR>

Query Match 55.8%; Score 2257; DB 2; Length 951;
Best Local Similarity 53.6%; Pred. No. 4.9e-147;
Matches 400; Conservative 130; Mismatches 176; Indels 40; Gaps 10;

QY 1 FVSIPIRYVETLVVADESMVKFHGADLEHYLLTLLATRAARLYRHPSILNPINIVVVKVLL 60

Db 237 FVSSPIRYVETLVVADESMVKFHGADLEHYLLTLLATRAARLYRHPSILNPINIVVVKVLLVI 296

QY 61 RDRSGPKVTGNAALTLRNFCNAWKLNKVSCKHPEYWDTAITLFTQDLGATTCDTLGM 120

Db 297 YEEQKGPVTSNAALTLRNFCNAWKLNKVSCKHPEYWDTAITLFTQDLGATTCDTLGM 356

QY 121 ADVTGMDPKRSCVTEDDGLPSAFTAHGLGHVFNMDHNVKVEEFGKLRANHMSP 180

Db 357 ADVTGMDPKRSCVTEDDGLPSAFTAHGLGHVFNMDHNVKVEEFGKLRANHMSP 416

QY 181 TLTIQIDRANPWSACAAIITDFLDSGHGDCLLDQPSKPISLPEDLPASVYLSQOCELA 240

Db 417 MLSSLDLQSPWSPCSAYMVVTSFLDNGHGECLMDKPNQPIKLPDLPGTLTDANRCQQT 476

QY 241 GYSGKPCP-YMOYCYKLVCTGKAGKQWCVOTPHFPWADGTCGGKGLCKGACVERNLN 299

Db 477 GEESKPCPDAASTCTTLMCTGTSGGLLYCQTKHFPWADGTCGGKGLCKGACVERNLN 535

QY 300 KH---RVDGSWAKWDPYGCPTCGGGVQLARRQCTNPTPANGKGYCBGVKVRSCNLE 356

Db 536 KHFAFVHGSWGPWGPWGDCSCVTCGGGQVYTMRECDNPVKNKGYCBGVKVRSCNIE 595

QY 357 PCPSSASGSGSFREQCEAFNGYNHSTNRLTLAVANVPKYSVGSVPDKCKLIRANGTGY 416

Db	596	DCPDN-NGKTFREEQCAHNEFSKASFGNEPTVETPKYAGVSPKDRCKLTCEAKGIGYF	654
QY	417	YVLAPKVVDTGLSPDSTSVCGKCIKAGCDGNLGSKKRFRKCGVCGGDNKSCKKVTGL	476
Db	655	FVLQPKVVDGTPCSPDSTSVCGQCVKAGCDRIIDSKKFKDKCGVCGGNGSTCKKMSGI	714
QY	477	FTPMHGYNVFVAIPAGASSIDIRQGYKGLIGDDNYLALKNSOGKYLNGHVFVSAVER	536
Db	715	VTSTRPGYHDIVIPAGATKATIEVKNQRNGRNGSFLAIRAADGYIILNGFTLSTLQ	774
QY	537	DLVYKGLRYSGTGTAVESLQASRPILPTVEVLSVSGKMTPPRVRYSYFLPKPREDK	596
Db	775	DLTYKGVLRYSGSSAALERIRSFPLKEPTLIQVLMVGHALRPKIKFTYEMKKKTES--	832
QY	597	SSHPKDPGRPSVLHNSVLSNQVEQDDPPRPARWAGSNWGPCSASGSLQKRAYDCRG	656
Db	833	-----FNAIPTES-----EWIEEWGECSTKCGSGQWRRVYVQCRD	867
QY	657	SAGORTVPACDAH--RPVETQAGE-PCPTWELSAWSPCKSGRGRFRRSLKCVGHGG	713
Db	868	INGH---PASECAKEVPASTRPCADLPCEPHQVGDWSPCKTCGKGYKRTLKCVSHDG	924
QY	714	RLARDQCNLHRKPQE--LDFCVLRPC	738
Db	925	GVLSNESCDPLKPKHYIDFCTLTQC	950
RESULT 2			
T00355			
hypothetical protein KIAA0688 - human			
C:Species: Homo sapiens (man)			
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000			
C:Accession: T00355			
R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.			
DNA Res. 5, 169-176, 1998			
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete			
A:Reference number: Z14142; MUID: 98403880; PMID: 9734811			
A:Accession: T00355			
A:Status: preliminary;			
A:Molecule type: mRNA			
A:Residues: 1-837 <ISH>			
A:Cross-references: EMBL:AB014588; NID:g3327189; PIDN:BA31663.1; PID:g3327190			
A:Experimental source: brain			
C:Genetics:			
A:Gene: KIAA0688			
C:Superfamily: thrombospondin type 1 repeat homology			
F:515-575/Domain: thrombospondin type 1 repeat homology <THR>			
Query Match 43.6%; Score 1761; DB 2; Length 837;			
Best Local Similarity 53.1%; Pred. No. 4.8e-113;			
Matches 324; Conservative 91; Mismatches 181; Indels 14; Gaps 8;			
QY	1	FVSIPIRYVETLVVADESVMVKFGADLEHYLTLLATARLYRHPISILNPIVIVVVKVLLL	60
Db	213	FASLSRFEVTLVVAADKAAAFHAGLKKRYLTVAATAAKAKFKHSIRNPVSLVTRLVIL	272
QY	61	RDRSDGPKVTGNAALTIRNFCAWQKKLNKSDKHPYWDTAITLFTRODLGATTCDTLGM	120
Db	273	GSSESGPQVPSAAQTFLRSFCAMQRLNTPEDSDPDHEDTAITLFTRODLGCVSTCDTLGM	332
QY	121	ADVGMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKCEEVFGKL-RANHMS	179
Db	333	ADVGMCDPARSCAIVEDDGLSAFTAAHELGHVFNHLDHNSKPCISLNGPLSTSRHYMA	392
QY	180	PTLIOIDRANPWSACSAIITDFLDSGHGDCLLDQPSKPISLPEDLPASVTLSSQCCRLA	239
Db	393	PVMAHVDPPEPWPSCARIFITDFLNGVGHCLLDKPEAPLHLPTVTFPKDADQOCOLT	452
QY	240	FGVSGKPCPYM-QYCTKLWCTGKAKGQMVQCTRHFPWADGTSCGEGKLCCLKACVERINL	298
Db	453	FGPDSRHCPQLPPPCAAALWCWSHLNGHAMCQTKHSPWADGTPCGPAQAQCMGRCILHMDQL	512
QY	299	NKHRVD--GSWAKWDPIGPCSTCGGGVQLARRQCTNTPTPANGGKYCEGVRYKRSCLNE	356
RESULT 3			
T47158			
hypothetical protein DKFp762C1110.1 - human (fragment)			
C:Species: Homo sapiens (man)			
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000			
C:Accession: T47158			
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.			
submitted to the Protein Sequence Database, March 2000			
A:Reference number: Z24379			
A:Accession: T47158			
A:Status: preliminary			
A:Molecule type: mRNA			
A:Residues: 1-550 <AAA>			
A:Cross-references: EMBL:AL162080			
A:Experimental source: adult melanoma (MeWo cell line); clone DKFp762C1110			
C:Genetics:			
A:Note: DKFp762C1110.1			
Query Match 39.7%; Score 1604; DB 2; Length 550;			
Best Local Similarity 49.7%; Pred. No. 1.8e-102;			
Matches 284; Conservative 106; Mismatches 142; Indels 40; Gaps 11;			
QY	175	NHMSPTLIQIDRANPWSACSAIITDFLDSGHGDCLLDQPSKPISLPEDLPASVTLSSQ	234
Db	10	SHMASLNSLDHSQWSPSCSAYMITSFLDNGHGECLMDKPNQPIQLPGDLFGTSYDANR	69
QY	235	QCELATGVGSKPCP-YMQVCTKLWCTGKAGQMVQCTRHFPWADGTSCGEGKLCCLKGACV	293
Db	70	QCQFTFGEDSKHCPDAASTCTLWCTGSGVLVQCTKHFPWADGTSCGEGKWCINGKV	129
QY	294	ERNLNKH---RVDSGWAKWDPIGPCSTCGGGVQLARRQCTNTPTPANGGKYCEGVRYKY	350
Db	130	NKTD-RKHEDTPHSGSNGMWGPWGDCSRTCGGGVQYTMRECDNPVKNKGKCEGRVRY	188
QY	351	RSCNLEPCPSASGKSFREOQCAFNGYNHSTNRLTLAVAWVPKYSGVSPDRCKLICRA	410
Db	189	RSCNLEPCPDN-NGKTFREEQCAHNEFSKASFGSPGPAVEWIPKYAGVSPKDRCKLICQA	247
QY	411	NGGYFVVLAPKVVVDGTLCSPDSTSVCGQCKCIKAGCDGNLGSKKRFRKCGVCGGDNKSCKV	470
Db	248	KGIGYFVLQPKVVVDGTPCSPDSTSVCGQCVKAGCDRIIDSKKFKDKCGVCGGNGSTC	307
QY	471	KKVTGLFTKPMHGYNVFVAIPAGASSIDIRQGYKGLIGDDNYLALKNSOGKYLNGHFV	530
Db	308	KTSGSVTSKAPGVHDIITPTGATNIEVQRNQRNNGSFLAIIKAADGTYILNGDYT	367
QY	531	VSAVERDLVVKGLLRYSVGTGTAVESLQASRPILPTVEVLSVSGKMTPPRVRYSYFLPK	590
Db	368	LSTLEQDQIMYKVVVLRYSGSSAALERIRSFPLKEPTLIQVLTGVNALRPKIKTYTFVKK	427

QY 591 EPREDKSSHPKDRGPSVLHNSVLSNQVEQDDPRPARWVAGSWGSPCSASCGLQKR 650
Db 428 K-----KES-----FNAIPFTS-----AWVIEWGECSKSCGLGWQR 460
QY 651 AYDCRGSAGQRTVPACAAH--RPVETQACGE-PCPTWELSAWSPCSKSGRGFORRLK 707
Db 461 LVECRDINGQ---PASECAKEVKPASTRPCADHPCPQWLGWSESCSKTCGKYKRSLK 517
QY 708 CVHGGRLRLARDQCNLHRKPQB-LDFCVLRPC 738
Db 518 CLSHDGGVLSHESCDPLKKPKHFIDFCTMAEC 549

RESULT 4
T13171
hypothetical protein F25H8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21371; T24896
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19413
A:Accession: T21371
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WTL>
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone F25H8
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19949
A:Accession: T24896
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2165 <W2>
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone T13H10
C:Genetics:
A:Gene: CESP:F25H8.3
A:Map position: 4
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1; 81

Query Match 38.8%; Score 1567; DB 2; Length 2165;
Best Local Similarity 39.6%; Pred. No. 2.9e-99;
Matches 321; Conservative 113; Mismatches 283; Indels 94; Gaps 17;

QY 7 YVETLVADSMVKFHGADLEHYLLTLTATAARLYRHPSILNPINIVVVKVLLLRDRDSG 66
Db 281 YVEVLVADTKMYEYHGRSLEDYVLTFTVASTIYRHQSLRASINVVVVKVLLVLTENAG 340
QY 67 PKVTGNAALTNRNFCAMOKKLNKVSXDKHPEYWDTAILTTRDLCAT--TCDTLGMADVGT 125
Db 341 PRITQNAQTLDQFCRWQYYNDPDDSSVQHDVAILTRKIDCRSQGKCDTLGLAELGT 400
QY 126 MCDPRKSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEV-----FGKL 172
Db 401 MCDMOKSCAILEDNGLSAFTIAHELGHVFSIPHDDEKCSYTPVNVKVFQSKFDKT 460
QY 173 RAN----HMSPTLIQIDRANPWSACSAIITFDLDSGGH--DCLLDQSKPSLPE---- 223
Db 461 QQNRFHIMAPTLEYNTHPWNSPCAGMLERFLFNENRGQTQCLFDQPVERRYEDVFR 520
QY 224 DLPGASYTLSSQCELAFAFGSKPCPYMOYCTKLACTGKAKGMVCOIHFHFWADGTCGE 283
Db 521 DEFGKKYDAHQCKVFVGPASELCYPTMTCRLMLCATYFGSOMGCRTOHMPWADGTPCDE 580
QY 284 GK--LCLKGCACVERNLNKRHVDGSKWAKWDYPGCSRTCGGVQLARQCTNPTPANGGK 341
Db 581 SRSMFCHGACVRLAPESLTKIDGOWDRWSNGECSRTCGGVQKGLRDCDPSPRNGGK 640
QY 342 YCEGVVYVYRSCNLEPCPSASGKSFREEQCEAFNGYN-----HSTNRLTLAVAVWPKY 395
Db 395

Db 641 YCVQRRYRSCNTQECFWDT--QPYREVOCSENNKNDIGIQVASTN-----THWVPKY 693
QY 396 SGVSPRDKCLICRANGTYGYVLAPKVDGTLTCLSPDSTSVCVQGCIRKAGCDNGLSKK 455
Db 694 ANVAPNERCKLYCLSGSAAFYLLRKVDGTPCDNRNGDDICVAGACMPAGCDQLHSTL 753
QY 456 RFDKCGVCGGDNCKKVTGLFTKP-MHGFNFVVAIPAGASSDIDRQGYKGLIGDNYL 514
Db 754 RRDCKGVCDDSGSKVVKGTFTNEQGTFFGYNEVNMKIPAGSANDIDIRKGYNNKKEDNYL 813
QY 515 ALKNSQKYLNLNGHFVVSVAVERDLVVKSLRLYSGTGTAVESLQASRPILPTVFLSV 574
Db 814 SLRAANGEFLNGHFQVSLARQIAFDTVLEYSGSDAIIERINGTPIRSDIYVHLSV 873
QY 575 GKMTPPRVRYSFYLPKEPRE-----DKSS 598
Db 874 GS-HPPDISYEYMTAAVNAVIRPISSALYLWRVTDTWTECDRACRQSQSKLCLMDMT 932
QY 599 HPKDPGPSVLHN-----SVLSLSNQVEQDDPRPARWVAGSWGSPCSASCGLQKRAVDC 654
Db 933 HRQS-----HNRNCONVLKPKQATRMCNIDCSTRWITEDVSSCAKCGSKQRQVSC 985
QY 655 RGSAGQRTVPA---CDAHRPVETOACGPPCP--TWELSAWSPCSKSCG-RGFQRRSLK 707
Db 986 VKMEGDRQTPASEHLCDRNSKPSDIASCIYDCSGRKNWYGEWTSCTCGSNGKMRKSY 1045
QY 708 CVHGGRLRLARDQCNLHRKFOELDFCVLRPC 738
Db 1046 CVDSDNRVRVDESCLGREQEATERECNRIPC 1076

RESULT 5
T18517
procollagen N-endopeptidase (EC 3.4.24.14) I - bovine
N:Alternate names: procollagen N-proteinase
C:Species: Bos primigenius taurus (cattle)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18517
R:Colige, A.; Nussgens, B.V.; Lapiere, C.M.
submitted to the EMBL Data Library, February 1996
A:Description: Cloning of the cDNA of the bovine procollagen I N-proteinase.
A:Reference number: Z18941
A:Accession: T18517
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1205 <COL>
A:Cross-references: EMBL:X96389; NID:e990769; PID:e228215; PIDN:CAA65253.1
A:Experimental source: skin
C:Genetics:
A:Gene: PC I-NP
C:Function:
A:Description: catalyzes cleavage of the propeptides of type I and II collagens prior
C:Keywords: hydrolase; metalloproteinase

Query Match 25.9%; Score 1045.5; DB 2; Length 1205;
Best Local Similarity 33.7%; Pred. No. 9.6e-64;
Matches 260; Conservative 108; Mismatches 300; Indels 103; Gaps 31;

QY 8 VETLVADSMVKFHGAD-LEHYLLTLTATAARLYRHPSILNPINIVVVKVLLLRDRDSG 66
Db 262 IEVLGVDSDSVQFQGHTEHVQKYLTLMTNIVNEIYHDESLGAHINVLVRIILSYGKSM 321
QY 67 PKV-TGNAALTNRNFCAMOKKLNKVSXDKHPEYWDTAILTTRDLCATTCDTLGMADVGT 125
Db 322 SLIEIGNPSQSLNVCNWAYLOQRPDTHDEYHDHAIFTLRQDF-GPSGMQ--GYAPVTG 378
QY 126 MCDPRKSCSVIEDDGLPSAFTTAHELGHVFNMPHDNV-KVCEVFGKLNHMHMSPTLIQ 184
Db 379 MCHPVRSCTLNHEDGSSAFVVAHETGHVLGMEHDEGQGNRCGD---EVLRLGSIAPLVA 435
QY 185 IDRANPWSACSAIITFDLDSGGHCDLLDQP----SKPISLPEDLPQASVTLSSQCELAFG 241
Db 436 AFHRFHWSRCSQELSRYLHS--YDCLRDDPFTHDWP-ALPQ-LPGLHYSMNEQCRDFG 491

Dd 173 ADKVVDGTCDSKNDICVGDCECLPVGCDGKLGSLLFKCGCKDGGSTCKTTEGRFDE 232

Qy 480 P--MHGVNFVVAIPAGASSIDIRQGKYGLIGDDNYIALKNKSOGKYLNLNGHFVVYSAVERD 537
||: : ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

Dd 233 RNLSPGYHDIIKLPEGATNIQEAR-----KSTNNLALKNGSDHFLVNLGLIQ-VEKE 286

Qy 538 LVWKGSLRLYSGTGTAVESLOASRPILPLTVLYSVGKMTPPRVRYSFYLPKPREPRED-- 595
: ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||: | ||:

Dd 287 VEVGGTFIVYD--DAEPETLSAOGPLSEELTVALLFRKSGRDTAIKYEFSLPLEEEVDYM 344

Qy 596 -----KSSHPKDRGPSVLHNSVLISLNQVEQPDDR-----PAR 630
: :

Dd 345 YKFDNWFPCSYSCGKGVTNRNLYCIDGNKGRVEDDLCEENNATKPFEKSKSETVDCEAE 404

Qy 631 WVASGWGPCSASC-SGLQKRAYDC-RGSAGRTVPACD---AAHRPVETOACGE-PCPT 684
| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

Dd 405 WFTGDWESCSTCDDQQOQYVWYCHOVFANGRRVTVEDGNCTVERPPVKOTCNRFACPE 464

Qy 685 WELSAWSPCSKSCGRGFQRRLSKC---VHGGRLLARDQCNLHRKPQELD---FCVLRP 737
| ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

Dd 465 WQAGFWACSCKEKGDAFYRSVTCRSEKEGEGKLLAADACPADEQ-EKFDTERTCNLGP 523

Qy 738 C 738

Dd 524 C 524

RESULT 7

T34395

hypothetical protein C37C3.6b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 01-Dec-2000

R:Geisel, C.; Bradshaw, H.

C:Accession: T34395; T34394

A:Submitted to the EMBL Data Library, July 1996

A:Description: The sequence of C. elegans cosmid C37C3.

A:Reference number: Z21518

A:Accession: T34395

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-2167 <GEI>

A:Cross-references: EMBL:U64857; PIDN:AAC25868.1; GSPDB:GN00023; CESP:C37C3.6b

A:Experimental source: strain Bristol N2; clone C37C3

A:Accession: T34394

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1555, 'SKF' <GE2>

A:Cross-references: EMBL:U64857; PIDN:AAC25867.1; GSPDB:GN00023; CESP:C37C3.6a

A:Experimental source: strain Bristol N2; clone C37C3

C:Genetics:

A:Gene: CESP:C37C3.6b; CESP:C37C3.6a

A:Map position: 5

A:Introns: 32/3; 104/2; 156/2; 207/1; 459/2; 536/3; 577/2; 1105/3; 1367/1; 1438/1; 15

Query Match 16.8%; Score 681; DB 2: Length 2167;
Best Local Similarity 33.7%; Pred. No. 2e-38;
Matches 162; Conservative 63; Mismatches 176; Indels 80; Gaps 19;

[illegible]

QY 538 LVVKGSLRLYSGTAVESLOASRPLEPLTVESLVSKMTTPRVRYSFYLPKEPRD-- 595
Db 287 VEVGTFIFVD--DAEPETLQAQPLSEELTVALLFKGSRDRAIKYEFISPLEEVEDYM 344
QY 596 -----KSSHPKDRGSPSVLHNSVLSNOVEQDDRP-----PAR 630
Db 345 YKFDNWTPCSVCSCGSGVOTRNLICIDGKNKGRVEDDCEENNATKPEFKSCETVDCEAE 404
QY 631 WYAGSWGPCSASCG--SGLOKRAVDC--RGSAGQRTVPACD--AAHRPVEVQACGE--PCPT 684
Db 405 WFTGWESCSSTCGDQGOYRVYCHOVFANGRRVTVEDGNCVTPVVKOTCNRACPE 464
QY 685 WELSAWSPCSKSCGGRFRRSLKC-----VHGGRLLARDQCNLHRKQELD---FCVLRP 737
Db 465 WQAGPWSACSEKCGAFQYRSVTCRSEKEGEGKLLAADACPADEQ--EKFTDERTCNLGP 523
QY 738 C 738
Db 524 C 524

RESULT 8
T16892
hypothetical protein T19D2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16892
R:Bentley, D.
A:Description: The sequence of C. elegans cosmid T19D2.
A:Reference number: Z18599
A:Accession: T16892
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-860 <BEN>
A:Cross-references: EMBL:U42846; NID:g1125809; PID:g1125810; PIDN:AAA83600.1; CESP:T19D2
A:Gene: CESP:T19D2.1
A:Introns: 56/3; 96/1; 171/3; 199/3; 245/2; 293/2; 335/3; 395/3; 426/3; 484/2; 505/3; 58

Query Match 12.6%; Score 510.5; DB 2; Length 860;
Best Local Similarity 26.1%; Pred. No. 3.5e-27;
Matches 198; Conservative 92; Mismatches 31; Indels 159; Gaps 40;

QY 8 VETLVVADSGM---VKFEGADLEHYLLTLATAAR---LYRHPSTLNPINIVVVKVLL 59
Db 135 VELAVFADDAMWDHFKKMYGRAAEENMTFFIMAVVNNIDVLYTORLLOPRINIKIVYEI 194
QY 60 LRD-----RDSGPKVTGNAALTURNFCAWQKKLNKVS DKHPEYWDTAILFTQDLCGATTC 115
Db 195 LKNIPHLNMRKHSNGVDRLDADFQYQNEINPNADPRHWDHALLFGYDLHRNGVK 254
QY 116 DTLGMAVGTCTMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEEFGKLRAN 175
Db 255 TVAGYAPVKCMGSGVRSCTINEGLDFGVSFVVTHEMGHSLGMYHDGNECD-----LRC- 308
QY 176 HMSNPTLIQIDRANPWSACSAIIITDLDGSH-GD-----CLLD---QPSKPISLPE-D 224
Db 309 CIMSPSV--GSGKTHWSQCSNEMATFV--GHLGDDFRPNPCLQDASANEQRMVAFKESE 364
QY 225 LGCASVTLQCELAFAFGVSKPC-----PYMQYCTKLWCYTKGAKGOMVOCOTRFPWP 275
Db 365 PPQLETLDEQCEIFHG-----ECWKHELKDGOTMONTICOMVWC---GNGEGVIRTAH-PA 416
QY 276 ADGTCGEGKGLCKGACVERHNLNKHRYDGSWAKWD--PYGPCSRTCG-----GGVQLAR 328
Db 417 LEGTYCGFMICRQGCYGVSSQLMRVTV--GKWSWTDNRPAFTCGRCSQCEIRGQIRMR 475
QY 329 --RQCNTPTPANGKYCEBGRVVKYRSCNLEPCPSSASGKSPREOCEAFNINYNSTNRLT 386
Db 476 SIROCNPNPSSNGGAPCGDEARGVMVCHRDVC--NGDSIENYATRYCSSLRDLNPAIPNTI- 533

QY 387 LAVAWPKYSGVS---PRDKCLICRANGTGYFYVLAPKVVVDGTLGSPDSTSVCVQKCI 443
Db 534 -----LSGEGQFQEQAMCKIWLISGNTNRTVS--NEPDGAPCGP--QYCIKGEGR 582
QY 444 KAGCDNIGSKRRDKCGVCGDNKSKKVTGLFTKPMHGYNFVVAIPAGASSIDIRQSG 503
Db 583 PLLCGSTTTLAYSEAD--CPL-----SVLQTT--TTPMPHVHSV----- 617
QY 504 YKGLIGDDNYLALKNSQKY----LLN---GHFVVS--AVERDLVWKGSLRLYSG--TGTA 553
Db 618 -----DOFAGKTNPKYKEHKKTFFLNEWSGWSVSECVIYDCHTQGVKVRRECLAGVC 670
QY 554 VESLOASRPLEPLTVESLVSKMTTPR--VRYSFYLPKEPRDKSHPKDPRGSPVLHN 611
Db 671 AGALRERQPCTRPCTGSRPL--TTSPPQOTFRNRFIAPLFPNRQT-----N 714
QY 612 SVLSLSNVEQPDPRPPARWAGSWGPCSCSGSLQ--KRAVDCRGSAGQRTVPACDAAH 670
Db 715 MTL-----RKVDHW--GPWSACSVTCGTGOKLRRENCIGQECATGTCV---- 757
QY 671 RPVETQACGEPCTPW--ELSAWSPCSKSCGGRFRRSLKC 708
Db 758 ----MQSCRENKNTWTEWSQSDCSVNGCGVQFRKRAC 792

RESULT 9
T18856
angiogenesis inhibitor homolog - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18856; T24653
R:McMurray, A.
A:Submitted to the EMBL Data Library, July 1995
A:Reference number: Z19031
A:Accession: T18856
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1444 <WT2>
A:Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN00028; CESP:C02B4.1
A:Experimental source: clone C02B4
R:McMurray, A.
A:Submitted to the EMBL Data Library, July 1995
A:Reference number: Z19917
A:Accession: T24653
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1444 <WT2>
A:Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:C02B4.1
A:Experimental source: clone T07C5
C:Genetics:
A:Gene: CESP:C02B4.1
A:Map position: X
A:Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3;

Query Match 12.5%; Score 504; DB 2; Length 1444;
Best Local Similarity 23.0%; Pred. No. 1.8e-26;
Matches 222; Conservative 91; Mismatches 306; Indels 348; Gaps 43;

QY 25 DLEHYL----LTLATAARLYRHPSTLNPINIVVVKVLLLRDRDS----GPKVTGNAALT 76
Db 239 DAEQHLLEFSLALINNVLVYQDDTLTPNLDIVIVRVEWRTQPSALSTGVHKNQQAQL 298
QY 77 LNFCAWQKKLNKVS--KHPEYWDTAILFTQDLCGATTCITLGMADVGTMCDPKRSV 135
Db 299 LDFCRYQAHMPGDLTDMNHYDHGVLITGYDIYHTTT--SVAGVAPVARMCDPLFACSL 357
QY 136 IEDDGLPSAFTTAHELGHVFNMPHDNVK--VCEEVEFGKLRANHMSPTLIQIDRANPWSAC 194
Db 358 VECGLHGRSFVLAHEMGHNMGMVHDGVQNCNKGCLMSAVNGAGKT-----TWSDC 409
QY 195 SAAITDFL----DSHGDCLLDQPSKPISLPE-----DLPGASYTLQOCELAFGVSK- 245
Db 410 SVREFNAFLQLQDESGRGNCLRDASPGLISTNHLSDRLPQGRFTADQOCSFYFGRDKY 469

F:46-106/Domain: thrombospondin type 1 repeat homology <THRL>

Query Match 12.2%; Score 493; DB 2; Length 951;
Best Local Similarity 22.2%; Pred. No. 6.3e-26;
Matches 171; Conservative 55; Mismatches 183; Indels 360; Gaps 27;

QY 305 GSAKWDPYGPGRSRTCGGQVQLARRC-----TNPTPANGKYCEGVVYKRSNLEPCPS 360
DB 51 GEMTKWTF--SRSCGGVTSQERHCLQORRKSVPGPNRTCTGTGSKRYQLCRVQEC- 106
QY 361 SASGKSFRECEQEARNG--YHNSHNLRLFLAVAWPKYSCVSPD-----KCLICRAN 411
DB 107 -PDGRSFRFEEQCSFNSHYNGRTHQ-----WKPLY-----PDDYVHSSSPCDLHC--TT 154
QY 412 GTGYFVLAPKVVDTGLGS-PDSTSVYQVKIKAGCDGNLGSKKRFRKCGVCGDNKSC 470
DB 155 VDGQQLMWP-ARDGTGSKLTLDRGVGVSGKEPFGDGLFSTHTLDKCGICQGDGSSC 213
QY 471 KKVTLGLTK-PMH-GYNFVVAIPAGASSIDIRQYKGLIGDDNYLALKNQSGYLLNGH 528
DB 214 THVTGNYRKGNAHLGYSLVTHIPAGARDIQIVERKKA-----DVLALADEAGYFFNGN 268
QY 529 FVYSAVERDLVYKSLRLYSGT-----GTAVESLQASRPILPTVEVLSV-GKMTPPRVR 583
DB 269 YKVDG--PKNFENTAGTVYKRRPMDVYETGIEYIVAGQPTNOGLNVVWVWQNGK--SPSIT 325
QY 584 YSFYLPKEPREDKS-----KD-----PR-----R 630
DB 326 FEYTLLOPHESRPQIIYYGFSESAESQGLDAGLMGFIPHNGSLYGOASSERLGLDNRL 385
QY 598 -SHP----- 600
DB 386 FGHFGLDMELPGSQGTNEVCEQAGGACGEGPRGKFRDRNVTGTLTGDKDDEEVD 445
QY 601 -----KD-----PR----- 604
DB 446 HFASQEFFSANISDOLLGAGSDLDKFTLNETVNSIFAQAPRSSILAESFFVDYENEGA 505
QY 605 GPSVLHNSVLSL-----NOVEQPDPRPPA----- 629
DB 506 GYLLNGSVLELSSDRVANSSEAPFPNVYSTLLTSAGNRTHKARTPRKARKOGVSPADM 565
QY 630 -----R 630
DB 566 YRWKLSHSPCSATCTGWSAYAMCVRDGVVEVDDSYCDALTRPEPVHFCAGRECQPR 625
QY 631 WVASGWSGPCSASCSGLQKRAV----- 552
DB 626 WETSSWSECSRTCGEGYQFRRVRCWKMLSPGFDSSVYSDICEAAEAARPEERKTCRNPAC 685
QY 653 -----DCRGSAGQRTV-----PACDAARHPVETOAC-GEPCP-TWELSA 689
DB 686 GPQWEMSESECTAKGERSVVTIRDCSDEKLCDPNTRPVGEKNCCTGPPCDRQWTVSD 745
QY 690 WSPCKSCGGRFQRRSLKCVHGGRLRLARDQCNLHRKQPQLDFCVLRPC 738
DB 746 WGPCSGSGGGRIRHRYCKTSDGRVVPESQCMETKPLAIHPCGDKNC 794

RESULT 11
T15976
hypothetical protein F08C6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15976
R:Bentley, D.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid F08C6.
A:Reference number: Z18440
A:Accession: T15976
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-957 <BEN>

QY 246 PCP-----YMOYCYLWCTGAKAGWMCQTRHFPWADTSCGEGKLCILKGACVE-RHNLNK 300
DB 470 EIPNGKAMDICRILWCGNSGS---TISTAH-PALEGSWCGANKWCHKGCQTHWTFGLTP 525
QY 301 HRVDGSAKWD-----PYGPCSRTCGGQVQLARRCCTNPTPANGKYCEGVVYKRSN 354
DB 526 VPIDGENSEGGAEGKCPIOQCAVSGSITVQGHRCVNPAPNNGGKTCGANIRGIVCG 585
QY 355 LEPCCPSASGKSFRFEO-----CEAF-----NGYHNSHNLRLFLAVAWPKYS 396
DB 586 ---ATSSNCLGFTREFEGNKICSSIKYDPHKPDQQLTGEGFEHST----- 627
QY 397 GVSPRDKCLICRANGT-----GYFVVLAPKVVDTGLCSFDPSTSVYQVKIKAGCDGN 450
DB 628 ---QPCRVWCHLIGSELIRNKGOF-----PDGTFPGFD--AYCVGGQCLALSCDNK 673
QY 451 ----- 450
DB 674 ALVEQPEDCPRIEGRSVHWEWSWSECSVSCGLGGRVYRERKCSSGRKCGVSESRP 733
QY 451 -----LGSKRF-----DKCG----- 461
DB 734 CEGVLRDCEEPGENKWCSEKCALGVQKFRPCLTDQCSSKHLQBERPCDNEGCTW 793
QY 462 -----VCGG-----DNKSKKVTGLFTKPMHGYNEVVAIPAGA 494
DB 794 DEWSSCSGCGGGRYRIKCLDKDQGDGDLKESCN-----TOKCISOSGWDWLPSCV 847
QY 495 S---SIDIRQRYGKGLIGDDNYLALKNQSGYLLNGHVFVSAVERDLVYKSLRLYS--- 548
DB 848 SCGIGFQIRER-----LCDGELCATANKQAR-TCNQOQCPSAF-----SLSVWSEWG 893
QY 549 -----GTGTAVESLQASRPILPTVEVLSVGVKMTPPRVRYSF 586
DB 894 EWTTCSATCGEGLQSRSCRRSGCTEDDASQTRCVNGPCEHSEYLTWSEWTTTCETCSF 953
QY 587 YLPKE-----PREDKSSHPKDPGRPSVLHNSVLSLSNQVEQ 622
DB 954 DSRRIACDGTTCNQDKIDEETCDIACLRKHSFGFISPRPKLI-----TSNDLRK 1007
QY 623 PDRP-----PARWVAGSWGPCSCSGGLQKRAVDGRGSAQRTVPACDAARHPVE 674
DB 1008 AFRGRLPLPIESIHSEKW---SEWGPCSVTCGSGRRVTRGCOEA-----SCPEQH--IQ 1056
QY 675 TQACG-EPCPTWEL---SAWSPCKSCGR-GFQRRSLKCVHGGRLRLARDQCNLH---RK 726
DB 1057 TEECNLNSC--LELFINDWSKSCSKSGQDGIQTRQKLC-----LFNNAECSSYAESRR 1108
QY 727 PQELDFC 733
DB 1109 CKDLFSC 1115

RESULT 10
T00260
hypothetical protein KIAA0605 - human
C:Species: Homo sapiens (man)
C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00260
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14086; MUID:98290545; PMID:9628581
A:Accession: T00260
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-951 <NAG>
A:Cross-references: EMBL:AB011177; NID:g3043733; PIDN:BAZ5531.1; PID:g3043734
A:Experimental source: brain
A:Genetics:
A:Note: KIAA0605
C:Superfamily: thrombospondin type 1 repeat homology

A:Experimental source: venom
R:Palme, M.J.; Desmond, H.P.; Theakston, R.D.; Crampton, J.M.
J. Biol. Chem. 267, 22869-22876, 1992
A:Title: Purification, cloning, and molecular characterization of a high molecular weight
11y.
A:Reference number: A44463; MUID:93054601; PMID:1385408
A:Accession: A44463
A:Molecule type: mRNA
A:Residues: 1-23, 'Q', 25-92, 'G', 94-131, 'G', 133-169, 'Q', 171-571 <PA2>
A:CROSS-references: GB:X68251
A:Experimental source: venom gland
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIP:118104)
R:Fujimura, Y.; Tital, K.; Usami, Y.; Suzuki, M.; Oyama, R.; Matsui, T.; Fukui, H.; Sug
Biochemistry 30, 1957-1964, 1991
A:Title: Isolation and chemical characterization of two structurally and functionally di
A:Reference number: A37958; MUID:91129280; PMID:1993206
A:Accession: A37958
A:Molecule type: protein
A:Residues: 360-372, 'E', 374-378, 'X', 380-384 <FUJ>
A:Note: 361-Val was also found
C:Comment: Inhibits collagen- and ADP-induced platelet aggregation.
C:Superfamily: mouse meltrin alpha; disintegrin homology
C:Keywords: hydrolase; metalloproteinase; venom; zinc
F:360-571/Product: jararagin C #status experimental <MAT>
F:362-444/Domain: disintegrin homology <DIS>
F:295, 299, 305/Binding site: zinc (His) #status predicted
F:296/Active site: Glu #status predicted

Query Match 6.4%; Score 258.5; DB 2; Length 571;
Best Local Similarity 24.2%; Pred. No. 4.4e-10;
Matches 117; Conservative 62; Mismatches 183; Indels 121; Gaps 26;

QY 6 RVETLVVADESMVFHAGDLEHYLLTLATAARLYRHPSTLNPI-NIVVVKVLL--LRD 62
Db 159 KIIEFVVVDQTVTKNGDLD-----KIKARYELANIVNEIFRYLYMHVALGLEI 211
QY 63 RDSGPKVT--GNAULTLRNFCQAWKLNKVSXKHPYWDTAILFTRQDLCGATCDILGM 120
Db 212 WSGDKITVKPDVDTLNSFAEWKRTDLLTRKKH-----DNAQLLTADFNG-----PTIGY 263
QY 121 ADVGIMCDPKRSCSVIED---DGLPSAFTTAHELGHVFNPHDVKVCEEVGKLRAH- 176
Db 264 AVIGSMCHPKRSGIVQDDYSPNLVAVVIMAHMGHNIGIHDT-----GSCSCQDY 315
QY 177 --MMSPTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQP-----SKPISLPEDLPGA 228
Db 316 PCIMGPT-ISNPSKFFSCSYIQCWDFIMNHNPEIINEPLGTDIISPPVCGNEL-- 371
QY 229 SYTLSSQCELATFGVSGKPCPYQYCTKLWCTGKAKQWVCQTRHFPWADGTSCEGKLC 288
Db 372 --EVGEECD---CGTPENCQNECCDA-----ATCKLK-----SGSQCHGD-- 407
QY 289 KGACVERHNLNKHRYD--GSAWAKWDPYPCSRCTCGGVQLARRQCTNPTPANGGKYQEGV 346
Db 408 ---CCQCKFSKSGTECRASMCECDPAEHCQTQSS-----ECPADVFKNGQPCLD- 455
QY 347 RVKYSRNLPCP-----SSAGSKSFREOCEAFNGYHNSTNRLTLAVAWV 392
Db 456 --NYGYCYNGNCPIIMYHQCVALFGADVAEEDSCFDKQKGNYYGYCRKENGKIPCA-- 511
QY 393 PKYSGVSPRD-KC-KLICRANGTG-----YFVLAPK-----VVDGTLGSPDSTSVCGV 440
Db 512 -----PEDVKGRGLYCKDNDSFGQNNPKMFYSNDEHKHGMVLPQTKCA--DGKVCNSG 562
QY 441 KCI 443
Db 563 HCV 565

RESULT 14
T00026
brain-specific angiogenesis inhibitor 1 - human

N:Alternate names: BAIL protein
C:Species: Homo sapiens (man)
C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 12-Feb-1999
C:Accession: T00026
R:Nishimori, H.; Shiratsuchi, T.; Urano, T.; Kimura, Y.; Kiyono, K.; Tatsumi, K.; Yos
submitted to the EMBL Data Library, June 1997
A:Reference number: Z14064
A:Accession: T00026
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1584 <NIS>
A:CROSS-references: EMBL:AB005297; NID:d1175078; PID:d1024528
A:Experimental source: brain
C:Genetics:
A:Gene: GDB:BAIL
A:CROSS-references: GDB:9838088; OMIM:602682
A:Map position: 8q24-8q24
C:Superfamily: thrombospondin type 1 repeat homology
F:408-462/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 6.3%; Score 255.5; DB 2; Length 1584;
Best Local Similarity 22.0%; Pred. No. 2.3e-09;
Matches 120; Conservative 42; Mismatches 174; Indels 209; Gaps 26;

QY 238 LAFGVGSKPCPYQYCTKLWCTGKAKQWVCQTRHFPWADGTSCEGKLCLEGA----- 291
Db 197 LAGRSRSHPCGIMQ--TPCACLGGEAGGP-----AAGPLAPRGDVLDRDANAGGPE 245
QY 292 -CVERHNLN- -HRVDGSAWAKWDPYPCSRCTCGGVQLARRQCTNPTPANGKYSCEGRV 348
Db 246 NCLTSLTQDRGGHGTGKWLWLGECTRDGGLQTRTCL-PAPGVEGGCGEVLE 304
QY 349 KYRSCNLEPC-----PSASGSKSPRE-----BQCEAFNGYHNSTNRLTLAVAWVPK 394
Db 305 EGRQCNREACGPAGRTSSRSQSLRSTARRREELGDELQOF-GFPAPQGTGDPAAEEW--- 360
QY 395 YSGVSPRDKCLICRANGTGYFYVLAPKVVDGTLCSPDSTSVCGKCIKAGCDNLGSK 454
Db 361 ----SPWSVCSSTC---GEGW-----QTRTRFCVSSS-YSTQCSGGLREQ 397
QY 455 KRFDKCGVCGDNNKSKKVTGLTKPMHGNFVVAIPAGASSIDIRQGYKGLIGDDNYL 514
Db 398 RLCNNSAVC-----PVHG-----ANDEW- 415
QY 515 ALKNSQKQYLLNGHFVSAVERDLVWKSLLRYSGTGTAVERLQASRPILPLEPTVEVLSV 574
Db 416 -----SPWSLCSSTC-----SPWSLCSSTC 425
QY 575 GKMTPPRVRYFYLPKEPREDKSHPKDPRGSPVLSNLSNQVEQDDRP----- 627
Db 426 GRGFRDRTR-----TCRPPQFGG-----NPCBGEKQTKFCNIAL 460
QY 628 -PARWVAG-----SWGPCASCSGSLQKRAVDCRGSAGQRTVPACDAARHPVETQACG- 679
Db 461 CPGRAVDGNMNEWSWSACSASCSQGRQRTRECNBP-----SYGGAECQGHVWETRDCLF 516
QY 680 EPCPT---WEL-SAWSPCSKSGRQFRRSLKCVG--HGGRL--ARD---QCNLHRKPQ 728
Db 517 QCCPVDKQWAWASGWGSCSVTCGAGSORRERVCSPFFGGACGQPDQVEYRQCGTQRCPE 576
QY 729 ELDFC 733
Db 577 PHEIC 581

RESULT 15
JC5928
semaphorin F precursor - human
C:Species: Homo sapiens (man)
C:Date: 10-Apr-1998 #sequence_revision 08-May-1998 #text_change 17-Nov-2000
C:Accession: JC5928
R:Simmons, A.D.; Puschel, A.W.; McPherson, J.D.; Overhauser, J.; Lovett, M.
Biochem. Biophys. Res. Commun. 242, 685-691, 1998

Search completed: April 29, 2003, 17:13:47
Job time : 34.07 secs

PI Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
PI Payne V;
XX WPI: 2002-041502/05.
DR N-PSDB; AAS97182.
XX Novel protease polypeptide useful for screening for substances that may
PT be used to treat, e.g., cancers, immune-related diseases,
PT cardiovascular disease, migraine, pain, psychotic and inflammatory
PT disorders -
XX
PS Claim 28: Figure 2G: 232pp; English.
XX The invention relates to an isolated, enriched, or purified protease
CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
CC screen for substances (S) that may modulate its activity. Administering
CC S (which modulates protease activity in vitro) may be used to treat a
CC disease or disorder selected from cancers (e.g., of tissues, of blood or
CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,
CC brain, ovarian, bladder or kidney), immune-related diseases and
CC disorders, cardiovascular disease, brain or neuronal-associated diseases
CC (e.g., central or peripheral nervous system diseases, migraine, pain,
CC sexual dysfunction, mood disorders, attention disorders, cognition
CC disorders, hypotension, hypertension, psychotic disorders, neurological
CC disorders and dyskinesias), metabolic disorders and inflammatory
CC disorders. (I) may also be useful as a diagnostic tool for a disease or
CC disorder such as those above. AAU72876-AAU72910 represent human
CC protease amino acid sequences of the invention.
XX
SQ Sequence 928 AA;

Query Match 100.0%; Score 2042; DB 23; Length 928;
Best Local Similarity 100.0%; Pred. No. 1.2e-162;
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSTLNINIVVKVLL 60
DB 238 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSTLNINIVVKVLL 297
QY 61 RDRDSGPKVTGNAALTFRNFCWQKLNKVDKHPYWDTAILETRQDLCGATTCDTLGM 120
DB 298 RDRDSGPKVTGNAALTFRNFCWQKLNKVDKHPYWDTAILETRQDLCGATTCDTLGM 357
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 180
DB 358 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 417
QY 181 TLQIDRANPWSACSAIIITDFLDGSGDCLLDQPSKPISLPEDLPGASYTLSSQCELA 240
DB 418 TLQIDRANPWSACSAIIITDFLDGSGDCLLDQPSKPISLPEDLPGASYTLSSQCELA 477
QY 241 GVSGKPCPYMOYCTKLWCTGKAKGQWVCQTRHFPWADGTSCGEGKLCCLKGACVERHNLN 300
DB 478 GVSGKPCPYMOYCTKLWCTGKAKGQWVCQTRHFPWADGTSCGEGKLCCLKGACVERHNLN 537
QY 301 HRVDGSAKWDYPGCSRTCCGGVQLARRQCTNPTPANGGKYCEGVKRVKRSNLEPCPS 360
DB 538 HRVDGSAKWDYPGCSRTCCGGVQLARRQCTNPTPANGGKYCEGVKRVKRSNLEPCPS 597
QY 361 SASGKSFRREQ 371
DB 598 SASGKSFRREQ 608

RESULT 2
AAG62299
ID AAG62299 standard; protein; 950 AA.
XX
AC AAG62299;
XX
DT 23-AUG-2001 (first entry)
XX Human metalloprotease MDTs6 protein.

XX Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;
KW osteopathic; antiarthritic.
XX Homo sapiens.
XX WO200134785-A1.
XX 17-MAY-2001.
XX 10-NOV-2000; 2000WO-JP07917.
XX 11-NOV-1999; 99JP-0321740.
XX 16-MAY-2000; 2000JP-0144020.
XX (YAMA) YAMANOUCHI PHARM CO LTD.
XX (KAZU-) KAZUSA DNA RES INST.
XX Yamaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
XX WPI: 2001-343602/36.
XX N-PSDB; AAH41003.
XX Metalloprotease with aggrecanase activity for treating joint diseases
PT especially osteoarthritis -
XX
PS Claim 1; Page 56-60; 85pp; Japanese.
XX This invention relates to a metalloprotease with aggrecanase activity.
CC The invention includes protein and DNA sequences of the metalloprotease,
CC vectors containing the DNA, host cells transformed by the vectors, and
CC antibodies directed against the metalloprotease. The antibodies, protein
CC and DNA sequences can be used in the treatment and prevention of joint
CC diseases, particularly osteoarthritis. The treatment may result in
CC osteopathic and antiarthritic activity. The present sequence represents
CC the metalloprotease of the invention termed MDTs6.
XX
SQ Sequence 950 AA;

Query Match 100.0%; Score 2042; DB 22; Length 950;
Best Local Similarity 100.0%; Pred. No. 1.2e-162;
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSTLNINIVVKVLL 60
DB 213 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSTLNINIVVKVLL 272
QY 61 RDRDSGPKVTGNAALTFRNFCWQKLNKVDKHPYWDTAILETRQDLCGATTCDTLGM 120
DB 273 RDRDSGPKVTGNAALTFRNFCWQKLNKVDKHPYWDTAILETRQDLCGATTCDTLGM 332
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 180
DB 333 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 392
QY 181 TLQIDRANPWSACSAIIITDFLDGSGDCLLDQPSKPISLPEDLPGASYTLSSQCELA 240
DB 393 TLQIDRANPWSACSAIIITDFLDGSGDCLLDQPSKPISLPEDLPGASYTLSSQCELA 452
QY 241 GVSGKPCPYMOYCTKLWCTGKAKGQWVCQTRHFPWADGTSCGEGKLCCLKGACVERHNLN 300
DB 453 GVSGKPCPYMOYCTKLWCTGKAKGQWVCQTRHFPWADGTSCGEGKLCCLKGACVERHNLN 512
QY 301 HRVDGSAKWDYPGCSRTCCGGVQLARRQCTNPTPANGGKYCEGVKRVKRSNLEPCPS 360
DB 513 HRVDGSAKWDYPGCSRTCCGGVQLARRQCTNPTPANGGKYCEGVKRVKRSNLEPCPS 572
QY 361 SASGKSFRREQ 371
DB 573 SASGKSFRREQ 583
RESULT 3


```
AAE22541
ID AAE22541 standard; Protein; 950 AA.
XX
AC AAE22541;
XX
DT 26-JUL-2002 (first entry)
XX
DE Human protease #2.
XX
KW Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme.
XX
OS Homo sapiens.
XX
PN WO200226949-A2.
XX
PD 04-APR-2002.
XX
PF 27-SEP-2001; 2001WO-US30350.
XX
PR 29-SEP-2000; 2000US-236689P.
XX
PA (LEXI-) LEXICON GENETICS INC.
XX
PI Friddle CJ, Hilbun E;
XX
DR WPI; 2002-372123/40.
DR N-PSDB; AAD35569.
XX
XX
Novel nucleic acid encoding a human protease, useful as a hybridization
probe for screening libraries and assessing gene expression patterns -
Claim 6; Page 36-38; 41pp; English.
XX
The present sequence is novel human protein (NHP), human protease.
CC NHPs share structural similarity with animal proteases particularly
CC zinc metalloproteases. Sequences of the invention are useful in
CC therapeutic, diagnostic and pharmacogenomic applications. NHP
CC polynucleotides are used as hybridisation probes for screening
CC libraries and assessing gene expression patterns. They can also be
CC used for treating related biological disorders such as obesity, high
CC blood pressure, arthritis, connective tissue disorders and infertility.
CC They are also used in gene therapy.
XX
SQ Sequence 950 AA;
Query Match 100.0%; Score 2042; DB 23; Length 950;
Best Local Similarity 100.0%; Pred. No. 1-2e-162;
Matches 371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLLATAARLYRHPISILNPINIVVKVLLL 60
DB 213 FVSIPRYVETLVVADESVMKFGADLEHYLLTLLATAARLYRHPISILNPINIVVKVLLL 272
QY 61 RDRSGPKVTGNAALTNRNFCAMQKLNKVDKHPYWDTAIFLTRODLGATCTDGLM 120
DB 273 RDRSGPKVTGNAALTNRNFCAMQKLNKVDKHPYWDTAIFLTRODLGATCTDGLM 332
QY 121 ADVGTMCDPKRSCVIEDGLPSAFTTAHELGHVFNPHDNKVCVEEFGKLRANHMWSP 180
DB 333 ADVGTMCDPKRSCVIEDGLPSAFTTAHELGHVFNPHDNKVCVEEFGKLRANHMWSP 392
QY 181 TLTIQIDRANPWSACSAIITFDLDSGHGDCLLDDQPSKIPSLPDLPGASYTLSQOCELAF 240
DB 393 TLTIQIDRANPWSACSAIITFDLDSGHGDCLLDDQPSKIPSLPDLPGASYTLSQOCELAF 452
QY 241 GVSGKPCPYMQYCTKLWCTGKAGQMWQCTRHFPWADGTSCGEGKCLKGACVERHNLNK 300
DB 453 GVSGKPCPYMQYCTKLWCTGKAGQMWQCTRHFPWADGTSCGEGKCLKGACVERHNLNK 512
QY 301 HRVDGSKWAKNDPYGPCSRCTCGGVQLARRQCTNPTPANGKYCEGVVKYRSCNLEPCPS 360
DB 513 HRVDGSKWAKNDPYGPCSRCTCGGVQLARRQCTNPTPANGKYCEGVVKYRSCNLEPCPS 572
QY 361 SASGKSFREQ 371
DB 573 SASGKSFREQ 583
RESULT 4
AAU74751
ID AAU74751 standard; Protein; 952 AA.
XX
AC AAU74751;
XX
DT 09-APR-2002 (first entry)
XX
DE Human protease PRTS-11 protein sequence.
XX
KW Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
KW cell proliferative disorder; developmental disorder; epilepsy;
KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
KW reproductive disorder; endometriosis.
XX
OS Homo sapiens.
XX
PN WO200198468-A2.
XX
PD 27-DEC-2001.
XX
PF 13-JUN-2001; 2001WO-US19178.
XX
PR 16-JUN-2000; 2000US-212336P.
PR 22-JUN-2000; 2000US-213955P.
PR 29-JUN-2000; 2000US-215396P.
PR 07-JUL-2000; 2000US-216821P.
PR 14-JUL-2000; 2000US-216946P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
PI Wallia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;
PI Azimzal Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
PI Kallick DA;
XX
WPI; 2002-090437/12.
N-PSDB; ABK12894.
XX
XX
Twenty one human proteases (referred to as PRTS-1 to PRTS-21), useful
in the diagnosis, treatment and prevention of gastrointestinal (e.g.
gastritis), cardiovascular (e.g. atherosclerosis) and cell
proliferative (e.g. cancer) disorders -
Claim 1; Page 144-146; 177pp; English.
XX
The present invention relates to twenty one new human proteases,
referred to as PRTS-1 to PRTS-21. The PRTS polynucleotides and
polypeptides of the invention are useful in the diagnosis, treatment and
prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
myocardial infarction, autoimmune/inflammatory e.g. acquired
immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
proliferative e.g. cancer, developmental e.g. Duchenne and Becker
muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
epilepsy and Alzheimer's disease and reproductive e.g. infertility and
endometriosis disorders. Numerous other examples of each disorder are
given in the specification. The present protein sequence represents
the human protease PRTS-11 protein of the invention.
XX
SQ Sequence 952 AA;
Query Match 100.0%; Score 2042; DB 23; Length 952;
```


FT Misc-difference 308 /note= "xaa= any amino acid"

XX WO200053774-A2.

XX 14-SEP-2000.

XX 08-MAR-2000; 2000WO-US06237.

XX 08-MAR-1999; 99US-0264585.

XX (NEUR-) NEUROCRINE BIOSCIENCES INC.

XX Kelner GS, Clark M, Maki RA;

XX WPI; 2000-594326/56.

XX N-PSDB; AAA95831.

XX Polynucleotide encoding novel members of a disintegrin,

XX metalloproteinase and thrombospondin domain protein family used to

XX prevent and treat Alzheimer's disease, cancer and autoimmune diseases -

XX Claim 12; Fig 23; 129pp; English.

XX The present sequence is human metalloproteinase ADAMTS-5. The

XX ADAMTS family of proteins is closely related to the ADAM (A Disintegrin

XX and Metalloproteinase Domain) family. Members of the ADAMTS family

XX contain a thrombospondin domain in addition to the disintegrin and

XX metalloproteinase domains found in the ADAMTS. ADAMTS polypeptides are

XX useful for the manufacture of medicaments for treating conditions

XX associated with neuroinflammation and/or neurodegeneration, such as

XX Alzheimer's disease, Parkinson's disease and stroke. They are also

XX useful for treating conditions associated with cell proliferation, cell

XX migration, inflammation and/or angiogenesis, such as cancer, arthritis

XX and autoimmune diseases. They can be used to treat patients afflicted

XX with an invasive tumour, a brain tumour or brain injury.

XX Sequence 381 AA;

Query Match 86.5%; Score 1767; DB 21; Length 381;

Best Local Similarity 93.3%; Pred. No. 4.6e-140;

Matches 322; Conservative 3; Mismatches 18; Indels 2; Gaps 2;

QY 28 HYLTLATARLVEHPSILNPINIVVVKVLLLRDRSGPKVTGNAALTNRNFCWQKLL 87

DB 2 HYRARAARAG-IFKHPISILNPINIVVVKVLLLRDRSGPKVTGNAALTNRNFCWQKLL 60

QY 88 NKVSDKHPEYWDTAILFTRODLGATTCDTLGMADVGTMCDDPKRSCSVIEDDGLPSAFTT 147

DB 61 NKVSDKHPEYWDTAILFTRODLGATTCDTLGMADVGTMCDDPKRSCSVIEDDGLPSAFTT 120

QY 148 AHELGHVFNPHDNNVKVCEEVFGKLRANHMMSPTLIQIDRANPWSACSAAIITDFLDGSH 207

DB 121 AHELGHVFNPHDNNVKVCEEVFGKLRANHMMSPTLIQIDRANPWSACSAAIITDFLDGSH 180

QY 208 GDCLLDQPSKIPSLPELPGASYTLISQCELAFCVGGKPCPYMOYCTKLKCTGKAKQMV 267

DB 191 GDCLLDQPSKIPSLPELPGASYTLISQCELAFCVGGKPCPYMOYCTKLKCTGKAKQMV 240

QY 268 CQTRHFWADGTCGEGKCLKLGACVERHNLNKRVDSWAKWDPYPCSRCTGGGVOLA 327

DB 241 CQTRHFWADGTCGEGKCLKLGACVERHNLNKRVDSWAKWDPYPCSRCTGGGVOLA 300

QY 328 RROCTNPTP-ANGGKCEGVVRYRSCNLEPCPSSASGKSFREQ 371

DB 301 RROXHQXPPLPTGKCEGVVRYRSCNLEPCPSSASGKSFREQ 345

RESULT 7

AAW78435

ID AAW78435 standard; Protein; 727 AA.

XX

AC AAW78435;

XX 11-MAY-1999 (first entry)

XX Human ADAMTS-1 protein.

XX Metalloproteinase-disintegrin protein with a thrombospondin domain; ARDS;

XX ADAMTS-1; drug composition; foodstuff; leukocyte; thrombocyte; hepatitis;

XX blood; erythrocyte; inflammatory disease; rheumatoid arthritis; asthma;

XX nephritis; Crohn's disease; acute respiratory disease syndrome.

XX Homo sapiens.

XX WO9855643-A1.

XX 10-DEC-1998.

XX 03-JUN-1998; 98WO-JP02449.

XX 03-JUN-1997; 97JP-0160422.

XX (KURE) KUREHA CHEM IND CO LTD.

XX Hakozaki M, Hirose K, Inoguchi E, Ishida Y, Ishioka K;

XX Kuno K, Matsushima K;

XX WPI; 1999-070277/06.

XX N-PSDB; AAX17990.

XX Human metalloproteinase-disintegrin protein with thrombospondin

XX domain - useful as leukocyte and thrombocyte decreasing and

XX erythrocyte increasing agent

XX Claim 1; Page 51-52; 82pp; Japanese.

XX This sequence represents a novel human metalloproteinase-disintegrin

XX protein with a thrombospondin domain (ADAMTS-1). The protein may be used

XX in drug compositions and foodstuffs, as an agent for decreasing the

XX leukocyte and thrombocyte blood count and increasing the erythrocyte

XX blood count, e.g. for treatment of inflammatory diseases such as

XX rheumatoid arthritis, hepatitis, nephritis, Crohn's disease, asthma

XX and ARDS.

XX

SQ Sequence 727 AA;

Query Match 65.5%; Score 1337; DB 20; Length 727;

Best Local Similarity 63.5%; Pred. No. 1.3e-103;

Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;

QY 1 FVSIPIRYETLVVVADESVMKFGADLEHYLLTLATARLVEHPSILNPINIVVVKVLLL 60

DB 13 FVSSPIRYETMLVADQSMAEFHGSLKHYLLTLFVAARLYKHPISIRNSVSLVVKILVI 72

QY 61 RDRSGPKVTGNAALTNRNFCWQKLLNKVSDKHPEYWDTAILFTRODLGATTCDTLGM 120

DB 73 HDQKQPEVTSNAALTNRNFCWQKQHNPPSDRDAEHYDTAILFTRODLGATTCDTLGM 132

QY 121 ADVGTMCDDPKRSCSVIEDDGLPSAFTTAEHLGHVFNPHDNNVKVCEEVFGKLRANHMMS 180

DB 133 ADVGTVCDDSRSCSVIEDDGLQNAFTAEHLGHVFNPHDNNVKVCEEVFGKLRANHMMS 192

QY 181 TLTIQIDRANPWSACSAAIITDFLDGSHGCLLDQPSKIPSLPELPGASYTLISQCELA 240

DB 193 MSLNLDHSPWSPCSAYMITSLFDNGHGBCLMDKPNQPIQLPGLPTGLYDANRQCQFTF 252

QY 241 GYGSKPCP-YMOYCTKLKCTGKAKQMVQTRHFWADGTCGEGKCLKLGACVERHNLN 299

DB 253 GEDSKHCPDAASTCTLWCTGTSGGVLYVQCTKHFWADGTCGEGKCLKLGACVERHNLN 311

QY 300 KH---RVDGSAWAKWDPYPCSRCTGGGVOLAARROCTNPTPANGGKCEGVVRYRSCNLE 356

DB 312 KHEDTFPHGSGWGPWGDGCSRTCGGGVQYTWRECDNFPVKNKGKCEGVVRYRSCNLE 371

QY 357 PCPSSASGKSFREQ 371

```
Db 372 DCPDN-NGKTFREEQ 385
      || : ||: |||||
RESULT 8
AAB21265
ID AAB21265 standard; Protein; 896 AA.
XX
AC AAB21265;
XX
XX
DT 23-FEB-2001 (first entry)
DE
DE Mouse metalloproteinase ADAMTS-1.
XX
XX Mouse; ADAMTS-1; metalloproteinase; ADAM;
KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; neurotropic; neuroprotective; antiparkinsonian;
KW cerebroprotective; cytotatic; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.
XX
OS Mus musculus.
XX
XX WO200053774-A2.
XX
XX 14-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US06237.
XX
XX 08-MAR-1999; 99US-0264585.
XX
XX (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX
XX Kelner GS, Clark M, Maki RA;
XX
XX WPI; 2000-594326/36.
XX
XX Polynucleotide encoding novel members of a disintegrin,
XX metalloproteinase and thrombospondin domain protein family used to
XX prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX Disclosure; Fig 17; 129pp; English.
XX
XX The present sequence is mouse metalloproteinase ADAMTS-1. The
XX ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
XX and Metalloproteinase Domain) family. Members of the ADAMTS family
XX contain a thrombospondin domain in addition to the disintegrin and
XX metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
XX useful for the manufacture of medicaments for treating conditions
XX associated with neuroinflammation and/or neurodegeneration, such as
XX Alzheimer's disease, Parkinson's disease and stroke. They are also
XX useful for treating conditions associated with cell proliferation, cell
XX migration, inflammation and/or angiogenesis, such as cancer, arthritis
XX and autoimmune diseases. They can be used to treat patients afflicted
XX with an invasive tumour, a brain tumour or brain injury.
XX
XX Sequence 896 AA;
Query Match 65.5%; Score 1337; DB 21; Length 896;
Best Local Similarity 63.28; Pred. No. 1.7e-103;
Matches 237; Conservative 51; Mismatches 81; Indels 6; Gaps 4;
Oy 1 FVSPRYVETLVVADESVMKFGHADLEHLLTLAATARLYRHPISILNPINVVVKVLL 60
Db 237 FVSPRYVETMLVAQDSMDPFGSLGKHLTLFSAAREYKHPISIRNSISLVVVKILVI 296
Oy 61 RDRSGPKVTGNAALTNRNCAWQKLNKYSDKHPEVWDTAILETRDLCGATTCDFLGM 120
Db 237 YEEQKGPVETVNAALTNRNCAWQKLNKYSDKHPEVWDTAILETRDLCGATTCDFLGM 356
Oy 121 ADVGTMCPKSCSVIEDDGLPSAFTTAHELGHVFNPNKVKCEVFGKLRANHWMS 180
Db 357 ADVGTMCPKSCSVIEDDGLQAAFTTAHELGHVFNPNKVKCEVFGKLRANHWMS 416
```

CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
CC toxins), or infectious. They can also be used to treat inflammatory
CC conditions, both chronic and acute conditions. The products can also be
CC used for detection and diagnosis. AA32002 to AA32080, and AA49503 to
CC AA49511 represent sequences given in the exemplification of the present
CC invention.

AA	
SQ	Sequence 950 AA;
	Query Match 65.3%; Score 1334; DB 20; Length 950;
	Best Local Similarity 63.5%; Pred. No. 3.2e-103;
	Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;

[illegible]

RESULT 10
AAB73549
ID AAB73549 standard; Protein; 950 AA;

AC AAB73549;

07-AUG-2001 (first entry)

Human ADAM-type metalloprotease MPTS4, SEQ ID NO:4.

Human; MDS4; ADAM-type metalloprotease; drug screening;
A Disintegrin And Metalloprotease; cancer; arthritis.

OS Homo sapiens.

AA
PN JP2001017183-A.XX
PD
23-JAN-2001XX
PF
09-.III.-1999.XX
09-III-1990. 00TP-0106594[illegible]

XX
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855

DR N-PSDB; AAH20224.

PT A new metal protease and its preparation for use as an anti-cancer and
PT anti-arthritis therapeutic -
XX

PS Claim 1; Page 12-14; 22pp; Japanese.

XX

CC The invention relates to the novel human ADAM (A Disintegrin And

CC Metalloprotease)-type metalloproteases MDTs4 (AAB73549) and MDTs5

CC (AAB73550). The metalloproteases can be used for the treatment of

CC cancers and arthritis. The invention also relates to the genes encoding

CC MDTs4 and MDTs5, vectors and host cells containing the MDTs4 or MDTs5

CC genes, the recombinant production of MDTs4 and MDTs5, and antibody

CC specific for MDTs4 or MDTs5, and methods of screening for compounds

CC which modulate the activity of MDTs4 and/or MDTs5. The present

CC sequence represents human MDTs4.

XX	Sequence	950 AA;
SQ		

Query Match	65.3%	Score 1334;	DB 22;	Length 950;
Best Local Similarity	63.5%	Pred. No. 3.2e-103;		
Matches 238;	Conservative	50;	Mismatches 81;	Indels 6;
				Gaps 4;

[illegible]

RESULT 11

AAB50002
ID AAB50002 standard; Protein; 950 AA.

XX AAB50002:

XX
DT 19-MAR-2001 (first entry)

XX DE Human METH1

Human; METH1; metalloproteinase; thrombospondin; angiogenesis inhibition;
cancer therapy; benign tumour; ocular angiogenic disease;
rheumatoid arthritis; psoriasis; wound healing; endometriosis;
vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;
coronary collateral; cerebral collateral; arteriovenous malformation;
ischaemic limb angiogenesis; Oster-Weber syndrome;
plaque neovascularisation; telangiectasia; haemophilic joint;
angiofibroma; fibromuscular dysplasia; wound granulation;
Crohn's disease; atherosclerosis; birth control.

Homo sapiens.

XX
PN WO200071577-A1.

30-NOV-2000.

XX PF 25-MAY-2000; 2000WO-US14462.
XX PR 25-MAY-1999; 99US-0318208.
XX PR 20-JUL-1999; 99US-0144882.
XX PR 10-AUG-1999; 99US-0147823.
XX PR 13-AUG-1999; 99US-0373658.
XX PR 22-DEC-1999; 99US-0171503.
XX PR 22-FEB-2000; 2000US-0183792.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (SMIK) SMITHKLINE BEECHAM CORP.
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX PA (IRUE/) IRUELA-ARISPE L.
XX PA (HAST/) HASTINGS G A.
XX PA (RUBE/) RUBEN S M.
XX PA (JONA/) JONAK Z L.
XX PA (TRUL/) TRULLI S H.
XX PA (FORN/) FORNWALD J A.
XX PA (TERR/) TERRETT J A.
XX PI Iruela-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
XX PI Fornwald JA, Terrett JA;
XX XX
XX XX
XX DR WPI: 2001-025136/03.
XX DR N-PSDB; AAC90057.
XX XX
XX PT METH1 and METH2 polynucleotides and encoded polypeptides, used to
XX PT inhibit angiogenesis in the treatment of disorders such as cancer,
XX PT rheumatoid arthritis and psoriasis -
XX XX
XX PS Claim 15; Fig 1: 768pp; English.
XX XX
CC The present sequence is human METH1 (ME for metalloprotease and TH for
CC thrombospondin). METH1 can be used for inhibiting angiogenesis in an
CC individual, and for treating cancer, benign tumours, an ocular angiogenic
CC disease, rheumatoid arthritis, psoriasis, delayed wound healing, nonunion
CC endometritis, vasculogenesis, granulations, hypertrophic scars, nonunion
CC fractures, scleroderma, trachoma, vascular adhesions, myocardial
CC angiogenesis, coronary collaterals, cerebral collaterals, arteriovenous
CC malformations, ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
CC fibromuscular dysplasia, wound granulation, Crohn's disease or
CC atherosclerosis. METH1 can also be used in birth control. METH1 can also
CC be used in diagnostic methods for the prognosis of cancer.
XX XX
SQ Sequence 950 AA;
Query Match 65.3%; Score 1334; DB 22; Length 950;
Best Local Similarity 63.5%; Pred. No. 3.2e-103;
Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;
QY 1 FVSTPRYVETLVVADESVMVKFHGADLEHYLLTLATAARLYRHPSILNPINIVVKVLL 60
DB 236 FVSSHRYVETLVVADESVMVKFHGADLEHYLLTLATAARLYRHPSILNPINIVVKVLL 295
QY 61 RDRDSGPKVTGNAALTLRNFCANQKLNKVSQKHPEYWDTAILETRQDLGGATCTGLM 120
DB 296 HDEQKGPVETVNAALTLRNFCANQKLNKVSQKHPEYWDTAILETRQDLGGATCTGLM 355
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHNDVNVKVEVFGKLRANHMSP 180
DB 356 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHNDVNVKVEVFGKLRANHMSP 415
QY 181 TLIIQDRANPWSACSAIIITFDLSDHGDCCLLDQPSKIPSLBEDLPASAYTLISQOCELAF 240
DB 416 MSLNLDHSQWSPSPCSAYMITSFLDNHGHECLMDKPNQIPLGDLPGTSDYANRQOQTF 475
QY 241 GVSGKPCP-TWQYCTKLWCTGKAGQWVCOTRHPFWDGTSCEGKLCCLKGACVERHNLN 299
DB 476 GEDSKHCPDAASTCLWCTGTSGVLVVCQTQKHFPWADGTSCEGKWCINGKVCNKTRD-R 534
QY 300 KH---RVDGSAKWADPYGPCSRTCGGGVQLARRQCTNPTPANGKYGCEGVRVRYRSCNLE 596

DB 535 KHFDTPPHSGWGMGPWGDSCRTCGGGVQVYTMRECDNPVKNKGKYGKGRVRYRSCNLE 594
QY 357 PCPSSASGKSFREQ 371
DB 595 DCPDN-NGKTFREQ 608
RESULT 12
AAW80285
ID AAW80285 standard; Protein; 967 AA.
XX AC AAW80285;
XX DT 19-JAN-1999 (first entry)
XX DE Human integrin ligand polypeptide ITGL-TSP.
XX KW ITGL-TSP; integrin ligand; angiogenic disease; cancer; atherosclerosis;
XX KW chronic inflammatory disorder; rheumatoid arthritis; tissue remodeling;
XX KW macular degeneration; diabetic retinopathy; Alzheimer's disease; human;
XX KW restenosis.
XX OS Homo sapiens.
XX PN EP874050-A2.
XX PD 28-OCT-1998.
XX PF 27-JAN-1998; 98EP-0300575.
XX PR 24-APR-1997; 97US-0845496.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PA (SMIK) SMITHKLINE BEECHAM CORP.
XX PI (SMIK) SMITHKLINE BEECHAM PLC.
XX PI Fronwald JA, Hastings GA, Jonak ZL, Terrett JA;
XX PI Trulli SH;
XX DR WPI: 1998-544643/47.
XX DR N-PSDB; AAV66508.
XX PT DNA encoding integrin ligand polypeptide ITGL-TSP - used to treat
XX PT angiogenic diseases; restenosis, Alzheimer's disease and in tissue
XX PT remodeling
XX PS Claim 11; Pages 6-9; 24pp; English.
XX XX
CC This represents a human integrin ligand polypeptide ITGL-TSP. ITGL-TSP
CC polypeptides can be used in the treatment of angiogenic diseases such as
CC cancer, cancer metastasis, chronic inflammatory disorders, rheumatoid
CC arthritis, atherosclerosis, macular degeneration or diabetic retinopathy,
CC restenosis, Alzheimer's disease and tissue remodeling. They can be used
CC to treat a subject in need of enhanced activity or expression of the
CC ITGL-TSP polypeptide.
XX SQ Sequence 967 AA;
Query Match 65.3%; Score 1334; DB 19; Length 967;
Best Local Similarity 63.5%; Pred. No. 3.3e-103;
Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;
QY 1 FVSIERYVETLVVADESVMVKFHGADLEHYLLTLATAARLYRHPSILNPINIVVKVLL 60
DB 253 FVSSHRYVETLVVADESVMVKFHGADLEHYLLTLATAARLYRHPSILNPINIVVKVLL 312
QY 61 RDRDSGPKVTGNAALTLRNFCANQKLNKVSQKHPEYWDTAILETRQDLGGATCTGLM 120
DB 313 HDEQKGPVETVNAALTLRNFCANQKLNKVSQKHPEYWDTAILETRQDLGGATCTGLM 372
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHNDVNVKVEVFGKLRANHMSP 180

Db 373 ADVGTVCDPSCSVIEDDGLQAFTTAHELGHVFNMPHDDAKQOCASLNGVNDSHMMAS 432
 QY 181 TLQIDRANPWSACSAIITDFDSHGDCLLDQPSKDISLPEDLPQASVYTLSSQOCELAF 240
 Db 433 MLNLDHSPQWSPCSAYMITSFLDNGHGECLMDKPNQIQLPGDLPGTSDANRQCQTF 492
 QY 241 GVGSKPCP-YMOYCTKLWCTGKAKGOMVQOTRHFPPWADGTSCGEGKLCCLKACVERHNLN 299
 Db 493 GEDSKHCPDAASTCTLWCTGTSGVLVCQTKHFPWADGTSCGEGKWICNKCYNKTD-R 551
 QY 300 KH---RVDSWAKWDYPGCSRTCCGGVQLARRQCTNPTPANGKYCBGVVRYRSCNLE 356
 Db 552 KHEDTPFHGSGWGMGPWDCSRTCCGGVQYTMRECDNPVPKNGKCYCBGKVRVYRSCNLE 611
 QY 357 PCPSSASGKSFRREQ 371
 Db 612 DCPDN-NGKTFREEQ 625

RESULT 13
 AAY04142
 ID AAY04142 standard; Protein; 967 AA.
 AC AAY04142;
 DT 15-JUN-1999 (first entry)
 XX Human Tango-71 protein.
 DE Human; Tango-71; Tango-73; Tango-74; Tango-76; Tango-83; diagnosis;
 KW detection.
 XX Homo sapiens.
 OS WO9907850-A1.
 PN 18-FEB-1999.
 XX 06-AUG-1998; 98WO-US16502.
 PR 05-SEP-1997; 97US-0058108.
 PR 06-AUG-1997; 97US-0054966.
 XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 PA Goodearl ADJ, Holtzman DA;
 PI WPI; 1999-167426/14.
 DR N-PSDB; AAX19955.
 XX New TANGO polypeptides and nucleic acids encoding them - useful as
 PT diagnostic agents and for treating disorders caused by aberrant
 PT expression of TANGO
 XX Claim 8; Fig 1; 84pp; English.
 CC The present sequence represents human Tango-71. Tango polypeptides are
 CC useful for identifying compounds which bind the polypeptide via direct
 CC binding, competition binding assays or Tango-71, -73, -74, 76 or -83-
 CC mediated signal transduction. Tango polypeptides are also useful for
 CC identifying modulating compounds by determining effect on Tango activity.
 CC Tango polypeptides and nucleic acids are useful for diagnosing diseases
 CC related to aberrant expression of Tango, and Tango polypeptides are
 CC useful for raising antibodies which can be used in diagnostic assays for
 CC detection of Tango, and also for generating anti-idiotypic antibodies for
 CC prevention and protection.
 XX Sequence 967 AA;

Query Match 65.3%; Score 1334; DB 20; Length 967;
 Best Local Similarity 63.5%; Pred. NO. 3.3e-103;
 Matches 236; Conservative 50; Mismatches 81; Indels 6; Gaps 4;

QY 1 FVSIPRIVETLVVADESMWKFHGADLEHYLLTLATAARLYRHPISILNPINIVVVKVLL 60
 Db 253 FVSSHRVETMLVADQSMAEFHGSLKHYLLTLFSAARLYKHPISIRNSVSLVVKVILVI 312
 QY 61 RDRDSGPKVTGNAALTLRNFCAKQKLNKYSKHPEDWDFAILFTRODLGGATTCDTLGM 120
 Db 313 HDQKGEVTSNAALTURNFCNWKQHNPSPDRDAHYDFAILFTRODLGGSTQCDTLGM 372
 QY 121 ADVGTMCDPKRCSVIEDDGLPSAFTTAHELGHVFNMPHNDNVKVCBEVFGKLRANHMSP 180
 Db 373 ADVGTVCDPSRCSVIEDDGLQAFTTAHELGHVFNMPHDDAKQOCASLNGVNDSHMMAS 432
 QY 181 TLQIDRANPWSACSAIITDFDSHGDCLLDQPSKDISLPEDLPQASVYTLSSQOCELAF 240
 Db 433 MLNLDHSPQWSPCSAYMITSFLDNGHGECLMDKPNQIQLPGDLPGTSDANRQCQTF 492
 QY 241 GVGSKPCP-YMOYCTKLWCTGKAKGOMVQOTRHFPPWADGTSCGEGKLCCLKACVERHNLN 299
 Db 493 GEDSKHCPDAASTCTLWCTGTSGVLVCQTKHFPWADGTSCGEGKWICNKCYNKTD-R 551
 QY 300 KH---RVDSWAKWDYPGCSRTCCGGVQLARRQCTNPTPANGKYCBGVVRYRSCNLE 356
 Db 552 KHEDTPFHGSGWGMGPWDCSRTCCGGVQYTMRECDNPVPKNGKCYCBGKVRVYRSCNLE 611
 QY 357 PCPSSASGKSFRREQ 371
 Db 612 DCPDN-NGKTFREEQ 625

RESULT 14
 AAB50011
 ID AAB50011 standard; Protein; 968 AA.
 AC AAB50011;
 DT 19-MAR-2001 (first entry)
 XX Protein; SEQ ID 125.
 DE Human; METH1; metalloprotease; thrombospondin; angiogenesis inhibition;
 KW cancer therapy; benign tumour; ocular angiogenic disease;
 KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
 KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
 KW scleroderma; trachoma; vascular adhesion; myocardial angiogenesis;
 KW coronary collateral; cerebral collateral; arteriovenous malformation;
 KW ischaemic limb angiogenesis; Osler-Webber syndrome;
 KW plaque neovascularisation; telangiectasia; haemophilic joint;
 KW angiofibroma; fibromuscular dysplasia; wound granulation;
 KW Crohn's disease; atherosclerosis; birth control.
 XX Homo sapiens.
 OS WO2000071577-A1.
 PN 30-NOV-2000.
 XX 25-MAY-2000; 2000WO-US14462.
 PF 25-MAY-1999; 99US-0318208.
 PR 20-JUL-1999; 99US-0144882.
 PR 10-AUG-1999; 99US-0147823.
 PR 13-AUG-1999; 99US-0373658.
 PR 22-DEC-1999; 99US-0171503.
 PR 22-FEB-2000; 2000US-0183792.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA (IRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBEN/) RUBEN S M.
 PA (JONAK/) JONAK Z L.
 PA (TRULLI/) TRULLI S H.

PA	(FORN/) FORNWALD J A.	XX	Human; secreted protein; fusion protein; gene therapy; protein therapy;
PA	(TERR/) TERRETT J A.	KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
XX		KW	developmental abnormality; foetal deficiency; blood; allergy; renal;
PI	Iruela-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;	KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
PI	Fornwald JA, Terrett JA;	KW	inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
XX		KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
DR	WPI; 2001-025136/03.	KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
XX		KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
PT	METH1 and METH2 polynucleotides and encoded polypeptides, used to	XX	
PT	inhibit angiogenesis in the treatment of disorders such as cancer,	OS	Homo sapiens.
PT	rheumatoid arthritis and psoriasis	XX	
PS	Claim 15; Pages 759-763; 768pp; English.	PH	Key
XX		FT	Misc-difference 40
XX	The present invention relates to human METH1 and METH2 (ME for	FT	/label= unknown
CC	metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).	FT	Misc-difference 45
CC	METH can be used for inhibiting angiogenesis in an individual, and for	FT	/label= unknown
CC	treating cancer, benign tumours, an ocular angiogenic disease,	FT	Misc-difference 169
CC	rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,	FT	/label= unknown
CC	vasculogenesis, granulations, hypertrophic scars, nonunion fractures,	FT	Misc-difference 293
CC	scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,	FT	/label= unknown
CC	coronary collaterals, cerebral collaterals, arteriovenous malformations,	FT	Misc-difference 297
CC	ischaemic limb angiogenesis, Osler-Webber syndrome, plaque	FT	/label= unknown
CC	neovascularisation, telangiectasia, haemophiliac joints, angiofibroma,	FT	Misc-difference 557
CC	fibromuscular dysplasia, wound granulation, Crohn's disease or	FT	/label= unknown
CC	atherosclerosis. METH can also be used in birth control. METH can also	XX	
CC	be used in diagnostic methods for the prognosis of cancer. The present	PN	WO9856804-A1.
CC	sequence is a protein isolated in the present invention.	XX	
XX		PD	17-DEC-1998.
SQ	Sequence 968 AA;	XX	
	Query Match 65.3%; Score 1334; DB 22; Length 968;	PF	11-JUN-1998; 98WO-US12125.
	Best Local Similarity 63.5%; Pred. No. 3.3e-103;	XX	
	Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;	PR	02-OCT-1997; 97US-0061060.
QY	1 FVSIPIRYVELTVADSWKFGHADLEHYLLTLLATAARLYRHPSTLNPINIVVVKVLL 60	PR	13-JUN-1997; 97US-0049547.
Db	254 FVSSHRVETMLVADQSMAEFFHSGSLKHLLTLFVAARLYRHPSTLRNSVSLVVKILVI 313	PR	13-JUN-1997; 97US-0049548.
QY	61 RDRDSGPKVTNAALTRNFCANVKLNKVDKHPYWDTAILTRQDLGGATTCTDGLM 120	PR	13-JUN-1997; 97US-0049549.
Db	314 HDEQKGEVTSNAALTRNFCANVKLNKVDKHPYWDTAILTRQDLGGATTCTDGLM 373	PR	13-JUN-1997; 97US-0049550.
QY	121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHNDNVKVEEVFGKLRANHMSP 180	PR	13-JUN-1997; 97US-0049606.
Db	374 ADVGTCDPKSCSVIEDDGLQAAFTTAHELGHVFNPHNDNVKVEEVFGKLRANHMSP 433	PR	13-JUN-1997; 97US-0049607.
QY	181 TLIDIDRANPWSACSAAILITDLDSDGDCLLDQPSKIPSLPEDLPASVTLSSQCELEAF 240	PR	13-JUN-1997; 97US-0049608.
Db	434 MSLNLDHSQWSPCSAYMITSFLDNGHGECIMDKPONPIQLPGLPSTIDANRQCOFTF 493	PR	13-JUN-1997; 97US-0049610.
QY	241 GVGSKPCP-YMQYCTKLWCTGKARGQWVCQTRHFPWADGTSCGEGKLCCLKGACVERHNLN 299	PR	13-JUN-1997; 97US-0049611.
Db	494 GEDSKHCPDAASTCSTLWCTGTSGVLVCQTKEHPWADGTSCGEGKWCINGKVCNKTD-R 552	PR	13-JUN-1997; 97US-0050566.
QY	300 KH---RVDGSAKWDPYGPSCSRCTGGGVQARQCQTNPTPANGKVCYGVYVYRSCNLE 356	PR	13-JUN-1997; 97US-0050901.
Db	553 KHFTDPFHSGWGMGPWDCSRTCGGVQYTMRECDNFPVKNKGKVCYGVYVYRSCNLE 612	PR	13-JUN-1997; 97US-0052989.
QY	357 PCPSASCKSFREQ 371	PR	18-AUG-1997; 97US-0051919.
Db	613 DCPDN-NGKTFREQ 626	PR	12-SEP-1997; 97US-0055984.
	RESULT 15	PR	12-SEP-1997; 97US-0058665.
AAW78189		PR	12-SEP-1997; 97US-0058668.
AC	AAW78189 standard; Protein: 967 AA.	PR	12-SEP-1997; 97US-0058669.
XX		PR	12-SEP-1997; 97US-0058750.
XX		PR	12-SEP-1997; 97US-0058971.
XX		PR	12-SEP-1997; 97US-0058972.
DT	13-APR-1999 (first entry)	PR	12-SEP-1997; 97US-0058975.
XX		PR	02-OCT-1997; 97US-0060834.
XX	Human secreted protein encoded by gene 64 clone HOUQC17.	PR	02-OCT-1997; 97US-0060841.
		PR	02-OCT-1997; 97US-0060844.
		PR	02-OCT-1997; 97US-0060865.
		PR	02-OCT-1997; 97US-0061059.
		XX	
		PA	(HUMA-) HUMAN GENOME SCI INC.
		XX	
		PI	Brewer LA, Ebner R, Ferrie AM, Feng P, Greene JM, Lafleur DW;
		PI	Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Young P;
		PI	Yu GL;
		XX	
		XX	WPI: 1999-080881/07.
		DR	N-PSDB; AAX04374.
		DR	
		XX	New isolated human genes and the secreted polypeptides they encode -
		PT	useful for diagnosis and treatment of e.g. cancers, neurological
		PT	disorders, immune diseases, inflammation or blood disorders

XX Claim 11; Page 297-300; 380pp; English.

XX This sequence represents a secreted human protein encoded by the gene

CC clone detailed in the descriptor line. The gene can be used to generate

CC fusion proteins by linking to the gene to a human immunoglobulin Fc

CC portion (e.g. AAX04302) for increasing the stability of the fused

CC protein as compared to the human protein only.

CC The invention relates to 86 novel genes and their fragments (nucleic

CC acid sequences: AAX04311-X04410; amino acid sequences AAW78126-W78225)

CC which are useful for preventing, treating or ameliorating medical

CC conditions e.g. by protein or gene therapy. Also, pathological

CC conditions can be diagnosed by determining the amount of the new

CC polypeptides in a sample or by determining the presence of mutations in

CC the new polynucleotides. Specific uses are described for each of the 86

CC polynucleotides, based on which tissues they are most highly expressed in

CC (see AAX04311 for described uses).

XX

SQ Sequence 967 AA;

Query Match 64.7%; Score 1321; DB 20; Length 967;

Best Local Similarity 62.9%; Pred. No. 4e-102;

Matches 236; Conservative 50; Mismatches 83; Indels 6; Gaps 4;

QY 1 FVSIPIRYVETLVVADSMVKFHGADLEHYLLTLLATARLYRHPHSILNPINIVVVKVLL 60

DB ||| |||||:||||:||||:|||||:||||:||||:||||:||||:||||:||||: 60

253 FVSSHRYVETMLVADQSMAEFHGSGLKHLYLLTFSVAARLKKHPIRNSVSLVVKILVI 312

QY 61 RDRDSGPKVTGNAALTNRNCAWKKLNKSDKHPEYWDTAILFTRODLGATTCDTLGM 120

DB ||| |||||:|||||:||||:||||:||||:||||:||||:||||:||||:||||: 120

313 HDEQKGPVETSNAAALTNRNFCNWKQHPSPDRDAEHYDTAILFTRODLGSGTCDTLGM 372

QY 121 ADVGTMCDPKRSCSVIEDDGLPSAETTAHELGHVFNPHDNVKNVCEEVFGKLRANHMSP 180

DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 180

373 ADVGTVCDPKSCSVIEDDGLQAFTTAHELGHVFNPHDDAKQACASLNGVNDQSHMAS 432

QY 181 TLIIQIDRANPWSACSAIITDLDSGHGDCLLDQPKPISLPEDLPASGYTLQQCELAF 240

DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| 240

433 MSLNLDHSGPSPCSAYMITSLDNGHGECLMDKPNQPIQLPGDLPGTSYDANRQCQTF 492

QY 241 GVGSKPCP-YMQYCKLWCTGKAGQMVQCTRHFPWADGTSCGGKGLCKGACVERHNLN 299

DB || || || || || || || || || || || || || || || || || || || || || 299

493 GEDSKHCPDAASTCSTLWCTGTSGLVLCQTKHFPWADGTSCGGKGLCKGACVERHNLN 551

QY 300 KH---RVDSNAKWDPYGCSRTCGGVQLARROCTNPTPANGKYCEGVKYSNLE 356

DB || || || || || || || || || || || || || || || || || || || || || 356

552 KHFDTPFHGSGMGMGPGWDCSRTCGGVQVYTMRECDNPVPKNGKGYCEGKRVYRSCNLE 611

QY 357 PCPSSASGKSFREQ 371

DB || : ||:||||| 371

612 DCPDN-NGKTFREQ 625

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 6.01802 seconds
(without alignments)
1813.869 Million cell updates/sec

Title: US-10-009-332-l_COPY_213_583
Perfect score: 2042
Sequence: 1 FVSIPIRYVETLVVADESWMK.....SCNLEPCPSASGKSFREO 371

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2.6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2.6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2.6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2.6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2.6/ptodata/1/iaa/PCTRUS_COMB.pep.*
6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1334	65.3	967	4	US-09-130-491-2
2	1182.5	57.9	837	4	US-09-122-126B-2
3	1136	55.6	608	4	US-09-130-491-13
4	1128	55.2	905	4	US-09-369-364A-9
5	1112	54.5	930	4	US-09-369-364A-2
6	1110	54.4	930	4	US-09-122-126B-15
7	1030	50.4	1882	4	US-09-369-364A-13
8	1019.5	49.9	874	4	US-09-369-364A-15
9	814	39.3	997	4	US-09-369-364A-7
10	687	33.6	551	4	US-09-130-491-16
11	654	32.0	1224	4	US-09-930-872-4
12	646.5	31.7	245	4	US-09-369-364A-11
13	631.5	30.9	1081	4	US-09-369-364A-17
14	629	30.8	518	4	US-09-369-364A-22
15	625.5	30.6	1211	4	US-09-491-522-5
16	618.5	30.3	1205	4	US-09-491-522-11
17	603	29.5	859	4	US-09-369-364A-5
18	493.5	24.2	481	4	US-09-130-491-8
19	421.5	20.6	566	4	US-09-491-522-7
20	332	16.3	491	4	US-09-930-872-2
21	229.5	11.2	1170	1	US-08-313-288B-20
22	214	10.5	802	4	US-09-632-098-2
23	214	10.5	812	4	US-09-632-098-4
24	205.5	10.1	441	3	US-08-985-526-3
25	203.5	10.0	751	2	US-08-836-442-3
26	202.5	9.9	529	2	US-08-836-442-3
27	196.5	9.6	814	4	US-09-813-819-4

28	196.5	9.6	814	4	US-09-920-048-4	Sequence 4, Appli
29	196.5	9.6	855	4	US-09-813-819-2	Sequence 2, Appli
30	196.5	9.6	855	4	US-09-920-048-2	Sequence 2, Appli
31	193.5	9.5	239	5	PCT-US93-01652-1	Sequence 1, Appli
32	184	9.0	201	4	US-09-411-329C-1	Sequence 1, Appli
33	184	9.0	201	4	US-09-411-329C-1	Sequence 1, Appli
34	184	9.0	201	4	US-09-466-276-1	Sequence 1, Appli
35	184	9.0	201	4	US-09-411-329C-3	Sequence 3, Appli
36	184	9.0	462	4	US-09-411-329C-17	Sequence 17, Appli
37	183.5	9.0	202	4	US-09-411-329C-16	Sequence 16, Appli
38	183.5	9.0	203	4	US-09-411-329C-5	Sequence 5, Appli
39	183.5	9.0	203	4	US-09-411-329C-15	Sequence 15, Appli
40	183.5	9.0	203	4	US-09-411-329C-3	Sequence 3, Appli
41	183.5	9.0	203	4	US-09-466-276-3	Sequence 3, Appli
42	183.5	9.0	464	4	US-09-411-329C-14	Sequence 14, Appli
43	183	9.0	462	4	US-09-026-001A-16	Sequence 16, Appli
44	179	8.8	592	4	US-09-026-001A-14	Sequence 14, Appli
45	178.5	8.7	200	2	US-08-836-442-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-130-491-2
; Sequence 2, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; EARLIER FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-130-491-2

Query Match	65.3%	Score	1334	DB	4	Length	967
Best Local Similarity	63.5%	Pred. No.	6.2e-113				
Matches	238	Conservative	50	Mismatches	81	Indels	6
Gaps	4						
Qy	1	FVSIPIRYVETLVVADESWMKFGADLEHYLLTLATAARLYRHPSTLNPINVVVKVLL	60				
Db	253	FVSSHRVETMLVADSMARFSGSLKHYLLTLFVAARLYRHPSTLNPINVVVKVLL	312				
Qy	61	RDRSGKPVGTGNAALTURNCAMQKLNKVSQKHPYWDFTAILFTRODLGGATTCDTLGM	120				
Db	313	HDEQKGPVTSNAALTURNCAMQKLNKVSQKHPYWDFTAILFTRODLGGATTCDTLGM	372				
Qy	121	ADVTGMDKPRKCSVIEDDGLPSAFTAHGLHVFVPMHNDKVNKVEEYFGKLRANHMSP	180				
Db	373	ADVTGMDKPRKCSVIEDDGLPSAFTAHGLHVFVPMHNDKVNKVEEYFGKLRANHMSP	432				
Qy	181	TLIOIDRANPWSASAAITDLDLSDHGCCLLDQPSKPSLPLDLPASVYTLSSQCELA	240				
Db	433	MLSNLDHQPWSPCSAYMTSFLDNGHGECLMDKPNQIQLPDLPGTSDANRQCQTF	492				
Qy	241	GVGSKPCP-YMYCTKLWCTGKAKQGVQVCTRHFPWADTSCGEGKCLKACVERHNL	299				
Db	493	GEDSKHCPDAASTCTLWCTGTSGGLVLCQTKHFPWADTSCGEGKCLKACVERHNL	551				
Qy	300	KH---RYDVGWAKWDVPGCSRTCGGGVQLARQCTNPTPANGKYCEGVRYKRSCLN	356				
Db	552	KHFTPPHSGWGMGPMGDCSRTCGGGVQVYTMEDCDNPVKNKGKCEGVRYKRSCLN	611				

```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-130-491-13

Query Match          55.6%; Score 1136; DB 4; Length 608;
Best Local Similarity 63.0%; Pred. No. 3.5e-95;
Matches 201; Conservative 39; Mismatches 73; Indels 6; Gaps 4;

QY 57 VLLERDRSGPKVTGNAALTLRNFCAWOKLKNVSKDKHPEYWDTAILFTRODLGGATTCD 116
Db 4 ILVIHQEGKEPVTISNAALTLRNFCAWOKLKNVSKDKHPEYWDTAILFTRODLGGATTCD 63
QY 117 TLGMADVTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVEFGKLRANH 176
Db 64 TLGMADVTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVEFGKLRANH 123
QY 177 MMSPTLIQIDRANPWSACSAAITDFLDSHGDCLLDQPSKPISLPDLPCASVYTLQQC 236
Db 124 MMSPTLIQIDRANPWSACSAAITDFLDSHGDCLLDQPSKPISLPDLPCASVYTLQQC 183
QY 237 ELAFVGSGKPCP-YMQYCTKLWCTGKAGQMVQCOTRHPHWDGTSCEGKLCGLKACVER 295
Db 184 QFTGEDSKHCPDAASTCTLWCTGSGVLCVQTKHFPWADGTSCEGKLCGLKACVER 243
QY 296 HNLNKH---RVGGWAKWDPYPCSRCTCGGGVQLARRQCTNPTPANGGKYCEGVRYKYS 352
Db 244 TD-RKHFTDPFHSGWGMGMPGDCSRCTCGGGVQLARRQCTNPTPANGGKYCEGVRYKYS 302
QY 353 CNLEPCFSSASGKSFREQ 371
Db 303 CNLEPCFSSASGKSFREQ 320

RESULT 4
US-09-369-364A-9
; Sequence 9, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 905
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-8
US-09-369-364A-9

Query Match          55.2%; Score 1128; DB 4; Length 905;
Best Local Similarity 53.8%; Pred. No. 3e-94;
Matches 205; Conservative 64; Mismatches 100; Indels 12; Gaps 6;

QY 1 FVSIPIRYVETLVVADESVMKFKHGADEHYLLTLATAARLYRHPISILNPINIVVKKVLL 60
Db 229 FVSEARFVETLVVADASMAAFYGTDLQNHILTVMSMAARIYKHPISIRNSVNLVVKVLL 288
QY 61 RDRDGPVKVTGNAALTLRNFCAWOKLKNVSKDKHPEYWDTAILFTRODLGGATTCD 119
Db 289 EKERWGPVSDNGGLTLRNFCSQWRFRNPKPSDRHPEHYDTAILFTRQNGPGEQCDTLG 348
QY 120 MADVGTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVEFGKLRANHMS 179
Db 349 MADVGTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVEFGKLRANHMA 408
QY 180 PTLIQIDRANPWSACSAAITDFLDSHGDCLLDQPSKPISLPDLPGAS--YTLQQC 237
```

```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-130-491-13

Query Match          55.6%; Score 1136; DB 4; Length 608;
Best Local Similarity 63.0%; Pred. No. 3.5e-95;
Matches 201; Conservative 39; Mismatches 73; Indels 6; Gaps 4;

QY 57 VLLERDRSGPKVTGNAALTLRNFCAWOKLKNVSKDKHPEYWDTAILFTRODLGGATTCD 116
Db 4 ILVIHQEGKEPVTISNAALTLRNFCAWOKLKNVSKDKHPEYWDTAILFTRODLGGATTCD 63
QY 117 TLGMADVTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVEFGKLRANH 176
Db 64 TLGMADVTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVEFGKLRANH 123
QY 177 MMSPTLIQIDRANPWSACSAAITDFLDSHGDCLLDQPSKPISLPDLPCASVYTLQQC 236
Db 124 MMSPTLIQIDRANPWSACSAAITDFLDSHGDCLLDQPSKPISLPDLPCASVYTLQQC 183
QY 237 ELAFVGSGKPCP-YMQYCTKLWCTGKAGQMVQCOTRHPHWDGTSCEGKLCGLKACVER 295
Db 184 QFTGEDSKHCPDAASTCTLWCTGSGVLCVQTKHFPWADGTSCEGKLCGLKACVER 243
QY 296 HNLNKH---RVGGWAKWDPYPCSRCTCGGGVQLARRQCTNPTPANGGKYCEGVRYKYS 352
Db 244 TD-RKHFTDPFHSGWGMGMPGDCSRCTCGGGVQLARRQCTNPTPANGGKYCEGVRYKYS 302
QY 353 CNLEPCFSSASGKSFREQ 371
Db 303 CNLEPCFSSASGKSFREQ 320

RESULT 4
US-09-369-364A-9
; Sequence 9, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 905
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-8
US-09-369-364A-9

Query Match          55.2%; Score 1128; DB 4; Length 905;
Best Local Similarity 53.8%; Pred. No. 3e-94;
Matches 205; Conservative 64; Mismatches 100; Indels 12; Gaps 6;

QY 1 FVSIPIRYVETLVVADESVMKFKHGADEHYLLTLATAARLYRHPISILNPINIVVKKVLL 60
Db 229 FVSEARFVETLVVADASMAAFYGTDLQNHILTVMSMAARIYKHPISIRNSVNLVVKVLL 288
QY 61 RDRDGPVKVTGNAALTLRNFCAWOKLKNVSKDKHPEYWDTAILFTRODLGGATTCD 119
Db 289 EKERWGPVSDNGGLTLRNFCSQWRFRNPKPSDRHPEHYDTAILFTRQNGPGEQCDTLG 348
QY 120 MADVGTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVEFGKLRANHMS 179
Db 349 MADVGTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNVKNVCEVEFGKLRANHMA 408
QY 180 PTLIQIDRANPWSACSAAITDFLDSHGDCLLDQPSKPISLPDLPGAS--YTLQQC 237
```

```
Db 409 PFFTHVNTLPWSPCSANVLTLLDGHGDCLLDAPTSLVPLTGLPCHSTLYELDQCK 468
QY 238 LATGVGSKPCP---YMQYCTKLWCTGKAGQWVCOTRH--FPWADGTSCEGKCLKLGAC 292
Db 469 QIFGPFHCPNTSVEDICVQJCAHRSDPEPCHTKNGSLWADGTPCGPGHLCJLDGSC 528
QY 293 VERHNLN--KHRVDGSAKWADPYGPCSRCTGGGVQLARRQCTNPTPANGKYCEGVRVKY 350
Db 529 VLKEDVENKAVVDGDPWGPWPWGQCSRTCGGGIQFSNRCDNPNQNGRFGCLGERVKY 588
QY 351 RSCNLEPCSSASGKSFREQ 371
Db 589 QSCNTEECF--PNGKSFREQ 607

RESULT 5
US-09-369-364A-2
; Sequence 2, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 930
; TYPE: PRT
; ORGANISM: mus musculus ADAMTS-5
US-09-369-364A-2

Query Match 54.5%; Score 1112; DB 4; Length 930;
Best Local Similarity 53.3%; Pred. No. 9e-93;
Matches 200; Conservative 59; Mismatches 106; Indels 10; Gaps 3;

QY 2 VSIPIRYVETLVWADSMVKFPGADLHLYLLTLATAARLYRHPSTILNPINIVVKKVLLR 61
Db 263 ISRAQVELLVADSSMARLYGRGLQHYLLTWASIANRLYSHASIHRLAVKVVVLT 322
QY 62 DRDSGPKVTGNAALTNRNCAWQKLNKYSDKHPEYWDTAIFTRQDLCGATTCDTLGM 121
Db 323 DKDTSLEYSKNAATTLKNFCKWQHQNGLDDEHEHYDAAILFTREDLCGHHSCTDLGMA 382
QY 122 DVGTMCDPKRCSCSVIEDDGLPSAFTAHGLHVNPHNDNVKVCVEFGKLRANHMSP 181
Db 383 DVGTICSPERSCAVIEDDGLHAFTVAHEIGHLLGLSHDDSKFCENFGTTEDKRLMSI 442
QY 182 LIQIDRANPWSACSAIITDLSHGDCCLLDQPSKPISLPDLPGASYTILSQOCEAF 241
Db 443 LTSIDASKPWSKCTSATITEFLDDHGNCLLDLPRKQILGPPELPQGYDATQCNLTFG 502
QY 242 VGSKPCPYMOYCTKLWCTGKAGQWVCOTRHFPWADGTSCEGKCLKLGACVER---H 296
Db 503 PEYSVCGMDVCAKLCVAVRQGMVCLTKLPVAVGTPCGKGRVLCQCKVDKTKKYY 562
QY 297 NLNKHVDGSAKWADPYGPCSRCTGGGVQLARRQCTNPTPANGKYCEGVRVKYRSCNLE 356
Db 563 STSSH--GNWGSMPWGQCSRSCGGGVQFAYRHNCNPAPRNNGRYCTGKRAIYRSCSV 619
QY 357 PCPSSASGKSFREQ 371
Db 620 PCP--PNGKSFREQ 632

RESULT 6
US-09-122-126B-15
; Sequence 15, Application US/09122126B
; Patent No. 6451575
```

```
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: AGGREGAN DEGRADING METALLO PROTEASES
; FILE REFERENCE: DM6909
; CURRENT APPLICATION NUMBER: US/09/122,126B
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 15
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-122-126B-15

Query Match 54.4%; Score 1110; DB 4; Length 930;
Best Local Similarity 53.6%; Pred. No. 1.4e-92;
Matches 201; Conservative 56; Mismatches 108; Indels 10; Gaps 3;

QY 2 VSIPIRYVETLVWADSMVKFPGADLHLYLLTLATAARLYRHPSTILNPINIVVKKVLLR 61
Db 263 ISRAQVELLVADSSMARLYGRGLQHYLLTWASIANRLYSHASIHRLAVKVVVLT 322
QY 62 DRDSGPKVTGNAALTNRNCAWQKLNKYSDKHPEYWDTAIFTRQDLCGATTCDTLGM 121
Db 323 DKDTSLEYSKNAATTLKNFCKWQHQNGLDDEHEHYDAAILFTREDLCGHHSCTDLGMA 382
QY 122 DVGTMCDPKRCSCSVIEDDGLPSAFTAHGLHVNPHNDNVKVCVEFGKLRANHMSP 181
Db 383 DVGTICSPERSCAVIEDDGLHAFTVAHEIGHLLGLSHDDSKFCENFGTTEDKRLMSI 442
QY 182 LIQIDRANPWSACSAIITDLSHGDCCLLDQPSKPISLPDLPGASYTILSQOCEAF 241
Db 443 LTSIDASKPWSKCTSATITEFLDDHGNCLLDLPRKQILGPPELPQGYDATQCNLTFG 502
QY 242 VGSKPCPYMOYCTKLWCTGKAGQWVCOTRHFPWADGTSCEGKCLKLGACVER---H 296
Db 503 PEYSVCGMDVCAKLCVAVRQGMVCLTKLPVAVGTPCGKGRVLCQCKVDKTKKYY 562
QY 297 NLNKHVDGSAKWADPYGPCSRCTGGGVQLARRQCTNPTPANGKYCEGVRVKYRSCNLE 356
Db 563 STSSH--GNWGSMPWGQCSRSCGGGVQFAYRHNCNPAPRNNGRYCTGKRAIYRSCSV 619
QY 357 PCPSSASGKSFREQ 371
Db 620 PCP--PNGKSFREQ 632

RESULT 7
US-09-369-364A-13
; Sequence 13, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1882
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-9
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (468)
; OTHER INFORMATION: Xaa = C
; NAME/KEY: MOD_RES
; LOCATION: (521)
; OTHER INFORMATION: Xaa = Y
US-09-369-364A-13
```

Query Match 50.4%; Score 1030; DB 4; Length 1882;
Best Local Similarity 51.3%; Pred. No. 6.2e-85;
Matches 193; Conservative 58; Mismatches 109; Indels 16; Gaps 8;

QY 1 FVSPRYVETLVADSWKFKGADLEHYLLTLATAARLYRHPSILNPINIVVKKVLL 60
DB 236 FLSYPRFVEVLVADNRVSYHGENLQHYLLTMSIVASYKDSIGNLINIVNLVI 295
QY 61 RDRDSGPKVTGNAALTLRNFCAMQKLNKVS DKHPEYWDTAILFTRODLGCA-TTCDTL 119
DB 296 HNEQDGPISFNAQTLLKNFCQWHSNSPGIHH---DTAVLLTRQDICRAHDKCDTL 351
QY 120 MADVGTMCOPKRSQSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVEFGKLRANHMS 179
DB 352 LAELGTICDPYRSCSISDLSGLSTAFTTAHELGHVFNPHDNVKNVCEVEFGKLRANHMS 410
QY 180 PTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKIPSLPEDLPFASYYTLSSQOCEL 238
DB 411 PTLNFTNPMWMSKRSKYIETFLDTGYGECLLNEPESRPYPLPVQLPGILYNVKNQXEL 470
QY 239 AFGVSGKPCPYMYCTKLWCT---TKAKGQMVQCOTRHPPWADGTSCGEGKCLKGACVER 295
DB 471 IFPGSQVCPYMQCRRLWCNNVNGVHKG---CRTQHTPWADGTCEPEGKCKXGFCVCPK 527
QY 296 HNLNKHVRDGSWAKWDPYGPCSRTCGGQVQLARROCTNPTPANGKGYCEGVVRYKRSCL 355
DB 528 -EMDVPVTDGWSGWSFGTSGRTCGGQVQLARROCTNPTPANGKGYCEGVVRYKRSCL 586
QY 356 EPCPSSASGKSPREQ 371
DB 587 EPCCLKQK--RDPREQ 600

RESULT 8
US-09-369-364A-15
; Sequence 15, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369, 364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 874
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-9
US-09-369-364A-15

Query Match 49.9%; Score 1019.5; DB 4; Length 874;
Best Local Similarity 50.9%; Pred. No. 2.1e-84;
Matches 192; Conservative 59; Mismatches 109; Indels 17; Gaps 9;

QY 1 FVSPRYVETLVADSWKFKGADLEHYLLTLATAARLYRHPSILNPINIVVKKVLL 60
DB 127 FLSYPRFVEVLVADNRVSYHGENLQHYLLTMSIVASYKDSIGNLINIVNLVI 186
QY 61 RDRDSGPKVTGNAALTLRNFCAMQKLNKVS DKHPEYWDTAILFTRODLGCA-TTCDTL 118
DB 187 HNEQDGPISFNAQTLLKNFCQWHSNSPGIHH---DTAVLLTRQDICRAHDKCDTL 242
QY 119 GMADVGTMCOPKRSQSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVEFGKLRANHMS 178
DB 243 GLAELGTICDPYRSCSISDLSGLSTAFTTAHELGHVFNPHDNVKNVCEVEFGKLRANHMS 301
QY 179 SPTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKIPSLPEDLPFASYYTLSSQOCEL 237
DB 302 APTLNFTNPMWMSKRSKYIETFLDTGYGECLLNEPESRPYPLPVQLPGILYNVKNQCE 361

QY 238 LAFGVSGKPCPYMYCTKLWCT---GKAKGQMVQCOTRHPPWADGTSCGEGKCLKGACVE 294
DB 362 LIFGPGSQVCPYMQCRRLWCNNVNGVHKG---CKTQHTPWADGTCEPEGKCKXGFCVCP 418
QY 295 RHLNKHVRDGSWAKWDPYGPCSRTCGGQVQLARROCTNPTPANGKGYCEGVVRYKRSCL 354
DB 419 K-EMEGPAIDGWSGWSFGTSGRTCGGQVQLARROCTNPTPANGKGYCEGVVRYKRSCL 477
QY 355 LEPSPSSASGKSPREQ 371
DB 478 TEPCKQK--RDPREQ 492

RESULT 9
US-09-369-364A-7
; Sequence 7, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369, 364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
US-09-369-364A-7

Query Match 39.9%; Score 814; DB 4; Length 997;
Best Local Similarity 46.8%; Pred. No. 1.2e-65;
Matches 176; Conservative 43; Mismatches 141; Indels 16; Gaps 10;

QY 2 VSIPRYVETLVADSWKFKGADLEHYLLTLATAARLYRHPSILNPINIVVKKVLL 60
DB 238 VSKKWCETLVADNRVSYHGENLQHYLLTMSIVASYKDSIGNLINIVNLVI 297
QY 61 RDRDSGPKVTGNAALTLRNFCAMQKLNKVS DKHPEYWDTAILFTRODLGCA-TTCDTL 118
DB 298 EDEEDLKIITHADNTLKSCKWQKSIKMGDAHPLHDTAILLTRKDLCAAMNRPCE 357
QY 119 GMADVGTMCOPKRSQSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVEFGKLRANH 177
DB 358 GLSHVAGMCQPHRSCSINEDTGLPFAFTVAHELGHVFNPHDNVKNVCEVEFGKLRANH 414
QY 178 MSPTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKIPSLPEDLPFASYYTLSSQOC 236
DB 415 MSPQLYDAAPLTWSRCSQYITRFLDRGWLCLDDPPAKDIIDFVSPVPGVLYDVSHQC 474
QY 237 ELAFGVSGKPCPYMYCTKLWCTGKAKGQMVQCOTRHPPWADGTSCGEGKCLKGACVER 295
DB 475 RLOYGAYSAFCEDMDNVCHTLWCSVGT---TCHSKLDAADVDTGRCGENKWLSCGCV 529
QY 296 HNLNKHVRDGSWAKWDPYGPCSRTCGGQVQLARROCTNPTPANGKGYCEGVVRYKRSCL 355
DB 530 VGFREPAVDGWSGWSAICSCSGMGVQSAERQCTQPTPKYKGYCVGERKFRCLNL 589
QY 356 EPCPSSASGKSPREQ 371
DB 590 QACP--AGRPSPRHVQ 603

RESULT 10
US-09-130-491-16
; Sequence 16, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.

; APPLICANT: Goodearl, Andrew D.J.
 ; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
 ; FILE REFERENCE: 09404/041001
 ; CURRENT APPLICATION NUMBER: US/09/130,491
 ; CURRENT FILING DATE: 1998-08-07
 ; EARLIER APPLICATION NUMBER: US 60/058,108
 ; EARLIER FILING DATE: 1997-09-05
 ; EARLIER APPLICATION NUMBER: US 60/054,961
 ; EARLIER FILING DATE: 1997-08-06
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 16
 ; LENGTH: 551
 ; TYPE: PRT
 ; ORGANISM: Rattus rattus
 ; US-09-130-491-16

Query Match 33.6%; Score 687; DB 4; Length 551;
 Best Local Similarity 56.4%; Pred. No. 2e-54;
 Matches 119; Conservative 29; Mismatches 57; Indels 6; Gaps 4;
 QY 165 CEVFGKLRANHMSTLQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPISLPED 224
 Db 1 CASLNGVSGDLSHLSMLSSLDHSPSPCSAYMVTSLFDNGHGCLMDKPNQIKLPSP 60
 QY 225 LPGAAYTLQOCELAFVGSKPCP-YMOYCTKLWCTGKAGKGMVQOTRHPFWADTSCGE 283
 Db 61 LFGTLDANRQCOFTFGESKHCPCDAASTCTTLWCTGSGGLLVQOTKHPFWADTSCGE 120
 QY 284 GKLCGLGACVERHNKHKH---RVDSWAKWDYPGCSRTCGGGVQLARRQCTNPTTPANGG 340
 Db 121 GKWCVSGKCVNKTDH-KHFAFVHSGWGPWGPWGDSCRTCGGGVYTMRECDNPVPKNGG 179
 QY 341 KYCEGVRYKRCNLPCPSSASGSKSFREQ 371
 Db 180 KYCEGKRVYRSCNIEDCPDN-NGKTFREQ 209

RESULT 11
 US-09-930-872-4
 ; Sequence 4, Application US/09930872
 ; Patent No. 6448388
 ; GENERAL INFORMATION:
 ; APPLICANT: Fride, Carl Johan
 ; APPLICANT: Hilbun, Erin
 ; TITLE OF INVENTION: No. 6448388 Human Proteases and Polynucleotides Encoding the Sa
 ; FILE REFERENCE: LEX-0219-USA
 ; CURRENT APPLICATION NUMBER: US/09/930,872
 ; CURRENT FILING DATE: 2001-08-14
 ; PRIOR APPLICATION NUMBER: US 60/225,852
 ; PRIOR FILING DATE: 2000-08-16
 ; NUMBER OF SEQ ID NOS: 5
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 1224
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 ; US-09-930-872-4

Query Match 32.0%; Score 654; DB 4; Length 1224;
 Best Local Similarity 39.5%; Pred. No. 5.6e-51;
 Matches 149; Conservative 49; Mismatches 149; Indels 30; Gaps 12;
 QY 8 VETLVVADSMVKFHG-ADLEHYLLTLATAARLYRHPHSILNPINIVVVKVLLLRDRDSG 66
 Db 292 VETLVVDDKMMQNHGHNITVYTLNMYSAFLKDGDTIGGNINIAIVGLILLEDEQPG 351
 QY 67 PKVTGNAALTLRNFCWAKKL-NKVSCKHPEYWDTAITLFTRODLGG--ATTCDTLGWADV 123
 Db 352 LVISHHADHTLSSFCQWGLMGKDGTRH---DHAILLTGLDICSWNKEPCDITLGFAP 407
 QY 124 GTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNV-KVCEVFVKLRANHMSP 182

Db 408 SGMCKYRSCNTINEDTGLGLAFTTAHESGHNFHMGHDEGNMCKKSEG-----NIMSPTL 462
 QY 183 IQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPI---SLPEDLPASVTLSSQOCELA 239
 Db 463 AGRNVFSWSPCSRQYLHKFLSTAQAICLADQP-KPVKEYKYPEKLPGLDYDANTQCKWQ 521
 QY 240 FGVSGKPCPY---MOYCTKLWC---TGRKAGQMVQOTRHPFWADTSCGEGKLCUKGACVE 294
 Db 522 FGEAKLCMLDFKKDICKALWCHRGRK-----CETKFMFAAEGTICGHDWCRGGQCVK 576
 QY 295 RHNLNKHRVDGWSWAKWDYPGCSRTCGGGVQLARRQCTNPTTPANGGYKCEGVRYKYRSCN 354
 Db 577 YGDEGPKPTHGWSWDSWSPCSRTCGGGVSHRSLCTNPKPSSHGGRFCEGSTRTLKLCN 636
 QY 355 LEPGPSSASGSKSFREQ 371
 Db 637 SQKCPRDS--VDFRAAQ 651

RESULT 12
 US-09-369-364A-11
 ; Sequence 11, Application US/09369364A
 ; Patent No. 6391610
 ; GENERAL INFORMATION:
 ; APPLICANT: Apte, Suneel
 ; APPLICANT: Hurskainen, Tiina L.
 ; APPLICANT: Hirohata, Satoshi
 ; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
 ; FILE REFERENCE: 26473/4007/10-30-00
 ; CURRENT APPLICATION NUMBER: US/09/369,364A
 ; CURRENT FILING DATE: 1999-08-06
 ; NUMBER OF SEQ ID NOS: 31
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 11
 ; LENGTH: 245
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens ADAMTS-8
 ; US-09-369-364A-11

Query Match 31.7%; Score 646.5; DB 4; Length 245;
 Best Local Similarity 57.9%; Pred. No. 3.5e-51;
 Matches 114; Conservative 41; Mismatches 41; Indels 1; Gaps 1;
 QY 1 FVSIPTRYVELTVVADESVMVPHGADLEHYLLTLATAARLYRHPHSILNPINIVVVKVLL 60
 Db 21 FVSEARFVETLLVADASMAAFYGADLQNLHILTLMSVAARIYKHPKSIKNSLNMVVKVLI 80
 QY 61 RDRDSGPKVTGNAALTLRNFCWAKKLNVKYSCKHPEYWDTAITLFTRODLGGAT--TCDTLG 119
 Db 81 EDEKGPESVNDGGLTLRNFCNMQRRFNQPSDRHPEHYDTAILLTRNFCGQGLCDTLG 140
 QY 120 MADVGTMCDPKRCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFVKLRANHM 179
 Db 141 VADIGTICDPKSCSVIEDGLQAAHTLAHELGHVLSMPHDDSKPCTRLFGPMGKHVMA 200
 QY 180 PTLIQIDRANPWSACSA 196
 Db 201 PLFVHLNQTLPWSPCSA 217

RESULT 13
 US-09-369-364A-17
 ; Sequence 17, Application US/09369364A
 ; Patent No. 6391610
 ; GENERAL INFORMATION:
 ; APPLICANT: Apte, Suneel
 ; APPLICANT: Hurskainen, Tiina L.
 ; APPLICANT: Hirohata, Satoshi
 ; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
 ; FILE REFERENCE: 26473/4007/10-30-00
 ; CURRENT APPLICATION NUMBER: US/09/369,364A
 ; CURRENT FILING DATE: 1999-08-06
 ; NUMBER OF SEQ ID NOS: 31


```

QY 188 ANPWSACSAIITDFLDGHDCLLDOP---SKPISLPEDLPASGYTLSSQOCELAFGVGS 244
Db 445 RPHWSKCSQOELSRYLHS--YDCLLDDPFAHDWP-ALPO-LFGLHYSMNEQCRFDGGLGY 500
QY 245 KPCPYMOY---CTKLWCTGKAKGOMVCOYTRHFPFADGTSCGEGKLCCLKGACV-----E 294
Db 501 MMCTAFRTDPCKQLWCS-HPDNPYFCKTKGPPLDGTMCAPGKHCFKGHCICWLTPTDILK 559
QY 295 RHNLNKHRYDGSWAKWDYPGCSRTCGGVQLARRQCTNPTPANGGKYCEGVVRKYRSCN 354
Db 560 R-----DGSWGAWSPEGSCSRTCGTGKTRQCONPHPANGGRTCSGLAYDFQLCS 611
QY 355 LEPCPSSASGKSFREQ 371
Db 612 RQDCPDSLA--DFREQ 626

```

Search completed: April 29, 2003, 17:11:47
Job time : 11.018 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:12:21 ; Search time 8.24692 Seconds

(without alignments)
3604.758 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_583

Perfect score: 2042

Sequence: 1 FVSIPRYVETLVVADESMVK.....SCNLEPCPSASGKSFREQ 371

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pap.*
- 2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pap.*
- 3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pap.*
- 4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pap.*
- 5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pap.*
- 6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pap.*
- 7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pap.*
- 8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pap.*
- 9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pap.*
- 10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pap.*
- 11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pap.*
- 12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pap.*
- 13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pap.*
- 14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2042	100.0	950	10	US-09-965-631-4
2	2020.5	98.9	823	9	US-10-163-316-2
3	1339	65.6	727	9	US-10-097-597-12
4	1339	65.6	727	9	US-10-097-580-12
5	1339	65.6	727	10	US-09-445-023A-12
6	1337	65.5	727	9	US-10-097-597-1
7	1337	65.5	727	9	US-10-097-580-1
8	1337	65.5	727	10	US-09-445-023A-1
9	1337	65.5	950	10	US-09-321-987B-4
10	1337	65.5	968	9	US-10-163-316-7
11	1334	65.3	967	12	US-10-105-929-2
12	1181.5	57.9	837	9	US-10-174-590-352
13	1181.5	57.9	837	9	US-10-176-758-352
14	1181.5	57.9	837	9	US-10-173-737-352
15	1181.5	57.9	837	9	US-10-173-706-352
16	1181.5	57.9	837	9	US-10-175-738-352
17	1181.5	57.9	837	9	US-10-175-752-352
18	1181.5	57.9	837	9	US-10-176-482-352
19	1181.5	57.9	837	9	US-10-176-757-352

20	1181.5	57.9	837	9	US-10-176-913-352	Sequence 352, App
21	1181.5	57.9	837	9	US-10-180-552-352	Sequence 352, App
22	1181.5	57.9	837	9	US-10-180-557-352	Sequence 352, App
23	1181.5	57.9	837	9	US-10-173-700-352	Sequence 352, App
24	1181.5	57.9	837	9	US-10-174-572-352	Sequence 352, App
25	1181.5	57.9	837	9	US-10-174-579-352	Sequence 352, App
26	1181.5	57.9	837	9	US-10-174-582-352	Sequence 352, App
27	1181.5	57.9	837	9	US-10-174-588-352	Sequence 352, App
28	1181.5	57.9	837	9	US-10-175-739-352	Sequence 352, App
29	1181.5	57.9	837	9	US-10-175-740-352	Sequence 352, App
30	1181.5	57.9	837	9	US-10-175-743-352	Sequence 352, App
31	1181.5	57.9	837	9	US-10-176-488-352	Sequence 352, App
32	1181.5	57.9	837	9	US-10-176-492-352	Sequence 352, App
33	1181.5	57.9	837	9	US-10-176-747-352	Sequence 352, App
34	1181.5	57.9	837	9	US-10-176-750-352	Sequence 352, App
35	1181.5	57.9	837	9	US-10-176-985-352	Sequence 352, App
36	1181.5	57.9	837	9	US-10-176-987-352	Sequence 352, App
37	1181.5	57.9	837	9	US-10-176-991-352	Sequence 352, App
38	1181.5	57.9	837	9	US-10-176-992-352	Sequence 352, App
39	1181.5	57.9	837	9	US-10-176-993-352	Sequence 352, App
40	1181.5	57.9	837	9	US-10-184-658-352	Sequence 352, App
41	1181.5	57.9	837	9	US-10-173-695-352	Sequence 352, App
42	1181.5	57.9	837	9	US-10-173-697-352	Sequence 352, App
43	1181.5	57.9	837	9	US-10-173-705-352	Sequence 352, App
44	1181.5	57.9	837	9	US-10-174-576-352	Sequence 352, App
45	1181.5	57.9	837	9	US-10-174-585-352	Sequence 352, App

ALIGNMENTS

RESULT 1
US-09-965-631-4
; Sequence 4, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Fridgde, Carl Johan
; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Enco
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 950
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-965-631-4

Query Match	100.0%	Score	2042	DB	10	Length	950
Best Local Similarity	100.0%	Pred. No.	1.8e-169				
Matches	371	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
QY	1	FVSIPRYVETLVVADESMVKFHGADLEHYLLTLATAARLYRHPSILNPNINIVVVKVLLL	60				
Db	213	FVSIPRYVETLVVADESMVKFHGADLEHYLLTLATAARLYRHPSILNPNINIVVVKVLLL	272				
QY	61	RDRDSGPKVTGNAALTLRNFCQWKKLNKVSQKHPEYWDTAILETRDGLCGATTCDTLGM	120				
Db	273	RDRDSGPKVTGNAALTLRNFCQWKKLNKVSQKHPEYWDTAILETRDGLCGATTCDTLGM	332				
QY	121	ADVGMCDPKRSCVIEDGLPSAFTTAHELGHVFNPHDNNKVCVEEFGKLRANHMSP	180				
Db	333	ADVGMCDPKRSCVIEDGLPSAFTTAHELGHVFNPHDNNKVCVEEFGKLRANHMSP	392				
QY	181	TLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPISLPDLPGASYTLSSOCELA	240				
Db	393	TLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPISLPDLPGASYTLSSOCELA	452				
QY	241	GVGSKPCPYMOCYTKLWCTGKAKGMVCQTRHFPWADGTSCGEGKCLKLGACVERHNLN	300				

```
Db 453 GVSGKPCPYMQYCTKLWCTGKAKGQVCOTRHFPPWADGTCGEGKCLKACVERHNLK 512
QY 301 HRVDGSKAKWDYPGCSRTCGGGVQLARRQCTNPTPANGKYCEGVYKYSRCSNLEPCPS 360
Db 513 HRVDGSKAKWDYPGCSRTCGGGVQLARRQCTNPTPANGKYCEGVYKYSRCSNLEPCPS 572
QY 361 SASGSKSFREQ 371
Db 573 SASGSKSFREQ 583

RESULT 2
US-10-163-316-2
; Sequence 2, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: A Human Matrix Metalloproteinase and Uses
; FILE REFERENCE: MP101-025P1RNM
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-316-2

Query Match 98.9%; Score 2020.5; DB 9; Length 823;
Best Local Similarity 94.2%; Pred. No. 1.1e-167;
Matches 371; Conservative 0; Mismatches 0; Indels 23; Gaps 1;

QY 1 FVSIPIRYVETLVVADESVMVKFPGADLEHYLLTLATARLYRHPSILNPINIVVKVLL 60
Db 213 FVSIPIRYVETLVVADESVMVKFPGADLEHYLLTLATARLYRHPSILNPINIVVKVLL 272
QY 61 RDRDSGPKVTGNAALTNRNFCAMQKLNKVS DKHPEYWDTAITLFTRODLCGATTCDTLGM 120
Db 273 RDRDSGPKVTGNAALTNRNFCAMQKLNKVS DKHPEYWDTAITLFTRODLCGATTCDTLGM 332
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRANHMSP 180
Db 333 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRANHMSP 392
QY 181 TLTIQIDRANPWSACSAAIITDFLDGSGHGDCLLDQPSKIPISLPDLPGASYTLSSQCELA 240
Db 393 TLTIQIDRANPWSACSAAIITDFLDGSGHGDCLLDQPSKIPISLPDLPGASYTLSSQCELA 452
QY 241 GVSGKPCPYMQYCTKLWCTGKAKGQVCOTRHFPPWADGTCGEGKCLKACVERHNLK 300
Db 453 GVSGKPCPYMQYCTKLWCTGKAKGQVCOTRHFPPWADGTCGEGKCLKACVERHNLK 512
QY 301 HR-----VDGSKAKWDYPGCSRTCGGGVQLARRQCTNPTPA 337
Db 513 HRPPDTIISPQKLLRLPNGLHTTQVDGSKAKWDYPGCSRTCGGGVQLARRQCTNPTPA 572
QY 338 NGKCYCEGVYKYSRCSNLEPCPSASGSKSFREQ 371
Db 573 NGKCYCEGVYKYSRCSNLEPCPSASGSKSFREQ 606

RESULT 3
US-10-097-597-12
; Sequence 12, Application US/10097597
; Publication No. US2003002352A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiiji
```

```
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: pharmaceutical
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 12
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-097-597-12

Query Match 65.6%; Score 1339; DB 9; Length 727;
Best Local Similarity 63.2%; Pred. No. 2e-108;
Matches 237; Conservative 52; Mismatches 80; Indels 6; Gaps 4;

QY 1 FVSIPIRYVETLVVADESVMVKFPGADLEHYLLTLATARLYRHPSILNPINIVVKVLL 60
Db 13 FVSIPIRYVETLVVADESVMVKFPGADLEHYLLTLATARLYRHPSILNPINIVVKVLL 72
QY 61 RDRDSGPKVTGNAALTNRNFCAMQKLNKVS DKHPEYWDTAITLFTRODLCGATTCDTLGM 120
Db 73 YEEQKGPVETVSNAAALTNRNFCAMQKLNKVS DKHPEYWDTAITLFTRODLCGATTCDTLGM 132
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRANHMSP 180
Db 133 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRANHMSP 192
QY 181 TLTIQIDRANPWSACSAAIITDFLDGSGHGDCLLDQPSKIPISLPDLPGASYTLSSQCELA 240
Db 193 MUSSLDHSQSPSCSAYMVTSLFNGHGECLMDKPNKIPKLPDLPGTLYDANRQCQFTF 252
QY 241 GVSGKPCPYMQYCTKLWCTGKAKGQVCOTRHFPPWADGTCGEGKCLKACVERHNLK 299
Db 253 GEESKHCPCDAASTCTTWTCTGSGLLVLCQTKHFPWADGTCGEGKWCVCVKNTDM- 311
QY 300 KH---RVDGSKAKWDYPGCSRTCGGGVQLARRQCTNPTPANGKYCEGVYKYSRCSNLE 356
Db 312 KHFAFPVHSGSPGMPGWDGCSRTCGGGVQYTMRECDNVPKNGKCYCEGKRVYRSCNIE 371
QY 357 PCPSSASGSKSFREQ 371
Db 372 DCPDN-NKGTFFREQ 385

RESULT 4
US-10-097-580-12
; Sequence 12, Application US/10097580
; Publication No. US20030032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiiji
; APPLICANT: Hakozaaki, Michinori
; APPLICANT: Ishioaka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; CURRENT FILING DATE: 2002-03-15
```

; PRIOR APPLICATION NUMBER: 09/445,023
 ; PRIOR FILING DATE: 1999-12-03
 ; PRIOR APPLICATION NUMBER: JP 9-160422
 ; PRIOR FILING DATE: 1997-06-03
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 727
 ; TYPE: PRT
 ; ORGANISM: Mus sp.
 US-10-097-580-12

Query Match 65.6%; Score 1339; DB 9; Length 727;
 Best Local Similarity 63.2%; Pred. No. 2e-108;
 Matches 237; Conservative 52; Mismatches 80; Indels 6; Gaps 4;

QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLTAARLYRHPHSILNPINIVVVKVLL 60
 DB 13 FVSSPRIVETMLVADQSMADFHGSLKHYLLTLFSAARFYKHPHSIRNSISLVVVKILVI 72
 QY 61 RDRSGPKVTGNAALTNRNFCAMOKKLNKYSVDKHPEYWDTAILFTRODLGATCTDTLGM 120
 DB 73 YEEQKGPEVTSNAALTNRNFCAMOKKLNKYSVDKHPEYWDTAILFTRODLGATCTDTLGM 132
 QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNKVCYEEVFGKLRANHMSP 180
 DB 133 ADVGTVCDDPSRSCSVIEDDGLQAFTTAHELGHVFNPHDDAKHACSLNGVSGDHLMAS 192
 QY 181 TLIQIDRANPWSACSAIITDLDLSDHGDCLLDQSPKPSLDPGLPGASYTLSSQCEIAF 240
 DB 193 MSLSDHSPWSPCSAYMVTSLDNGHGECLMDKPNPIKPSLDPLGTLYDANRQCQFTF 252
 QY 241 GVSGKPCP-YMQVCTKLWCTGKAKGVQVOTRHFPPWADGTSCGGVQLARRQCTNPTPANGKYCEGVRYKRSCLNLE 299
 DB 253 GEESKHCPPDAASTCTTLCCTGGLVCQTKHFPWADGTSCGGKWCVCVKNTKDM- 311
 QY 300 KH---RVDSGWAKWDYPCSRCTCGGVQLARRQCTNPTPANGKYCEGVRYKRSCLNLE 356
 DB 312 KHFPATPVHSGWGPWPGWDCSRTCTCGGVQVYTMRECDNPVPKNGGKYCEGKRVRYRSCNIE 371
 QY 357 PCPSSASGKSFRREQ 371
 DB 372 DCPDN-NGKTFREQ 385

RESULT 5
 US-09-445-023A-12
 ; Sequence 12, Application US/09445023A
 ; Patent No. US20020119167A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hirose, Kunitaka
 ; APPLICANT: Inoguchi, Eiji
 ; APPLICANT: Hakozaiki, Michinori
 ; APPLICANT: Ishioka, Keiko
 ; APPLICANT: Ishida, Yukako
 ; APPLICANT: Matsushima, Kouji
 ; APPLICANT: Kuno, Kouji
 ; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
 ; TITLE OF INVENTION: composition and method of immunologically analyzing human ADAMTS
 ; FILE REFERENCE: Q57092
 ; CURRENT APPLICATION NUMBER: US/09/445,023A
 ; CURRENT FILING DATE: 1999-12-03
 ; PRIOR APPLICATION NUMBER: JP 9-160422
 ; PRIOR FILING DATE: 1997-06-03
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 727
 ; TYPE: PRT
 ; ORGANISM: Mus sp.
 US-09-445-023A-12

Query Match 65.6%; Score 1339; DB 10; Length 727;

Best Local Similarity 63.2%; Pred. No. 2e-108;
 Matches 237; Conservative 52; Mismatches 80; Indels 6; Gaps 4;
 QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLTAARLYRHPHSILNPINIVVVKVLL 60
 DB 13 FVSSPRIVETMLVADQSMADFHGSLKHYLLTLFSAARFYKHPHSIRNSISLVVVKILVI 72
 QY 61 RDRSGPKVTGNAALTNRNFCAMOKKLNKYSVDKHPEYWDTAILFTRODLGATCTDTLGM 120
 DB 73 YEEQKGPEVTSNAALTNRNFCAMOKKLNKYSVDKHPEYWDTAILFTRODLGATCTDTLGM 132
 QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNKVCYEEVFGKLRANHMSP 180
 DB 133 ADVGTVCDDPSRSCSVIEDDGLQAFTTAHELGHVFNPHDDAKHACSLNGVSGDHLMAS 192
 QY 181 TLIQIDRANPWSACSAIITDLDLSDHGDCLLDQSPKPSLDPGLPGASYTLSSQCEIAF 240
 DB 193 MSLSDHSPWSPCSAYMVTSLDNGHGECLMDKPNPIKPSLDPLGTLYDANRQCQFTF 252
 QY 241 GVSGKPCP-YMQVCTKLWCTGKAKGVQVOTRHFPPWADGTSCGGVQLARRQCTNPTPANGKYCEGVRYKRSCLNLE 299
 DB 253 GEESKHCPPDAASTCTTLCCTGGLVCQTKHFPWADGTSCGGKWCVCVKNTKDM- 311
 QY 300 KH---RVDSGWAKWDYPCSRCTCGGVQLARRQCTNPTPANGKYCEGVRYKRSCLNLE 356
 DB 312 KHFPATPVHSGWGPWPGWDCSRTCTCGGVQVYTMRECDNPVPKNGGKYCEGKRVRYRSCNIE 371
 QY 357 PCPSSASGKSFRREQ 371
 DB 372 DCPDN-NGKTFREQ 385

RESULT 6
 US-10-097-597-1
 ; Sequence 1, Application US/10097597
 ; Publication No. US20030022352A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hirose, Kunitaka
 ; APPLICANT: Inoguchi, Eiji
 ; APPLICANT: Hakozaiki, Michinori
 ; APPLICANT: Ishioka, Keiko
 ; APPLICANT: Ishida, Yukako
 ; APPLICANT: Matsushima, Kouji
 ; APPLICANT: Kuno, Kouji
 ; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
 ; TITLE OF INVENTION: pharmaceutical
 ; TITLE OF INVENTION: composition and method of immunologically analyzing human AD
 ; FILE REFERENCE: Q57092
 ; CURRENT APPLICATION NUMBER: US/10/097,597
 ; CURRENT FILING DATE: 2002-03-15
 ; PRIOR APPLICATION NUMBER: 09/445,023
 ; PRIOR FILING DATE: 1999-12-03
 ; PRIOR APPLICATION NUMBER: JP 9-160422
 ; PRIOR FILING DATE: 1997-06-03
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 727
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-097-597-1

Query Match 65.5%; Score 1337; DB 9; Length 727;
 Best Local Similarity 63.5%; Pred. No. 3e-108;
 Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;

QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLTAARLYRHPHSILNPINIVVVKVLL 60
 DB 13 FVSSPRIVETMLVADQSMADFHGSLKHYLLTLFSAARLYRHPHSIRNSISLVVVKILVI 72
 QY 61 RDRSGPKVTGNAALTNRNFCAMOKKLNKYSVDKHPEYWDTAILFTRODLGATCTDTLGM 120
 DB 73 HDEQKGPEVTSNAALTNRNFCAMOKKLNKYSVDKHPEYWDTAILFTRODLGATCTDTLGM 132


```

; Sequence 4, Application US/09321987B
; Patent No. US20020102210A1
; GENERAL INFORMATION:
; APPLICANT: Kimble, Judith E
; APPLICANT: Blleloch, Robert H
; TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
; FILE REFERENCE: 960296 95386
; CURRENT APPLICATION NUMBER: US/09/321,987B
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087,170
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/129,023
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Murine
US-09-321-987B-4

Query Match      65.5%; Score 1337; DB 10; Length 950;
Best Local Similarity 63.2%; Pred. No. 4.2e-108;
Matches 237; Conservative 51; Mismatches 81; Indels 6; Gaps 4;

Qy 1 FVSPRYVETLWVADESVMKFGADLEHYLLTLATARLYRHPHSILNPINIVVVKVLL 60
Db 237 FVSPRYVETMLVADQSMADFHGSLKHYLLTLFVAAFYKHPHSIRNSISLVVVKILVI 296

Qy 61 RDRSGPKVTGNAALTLRNFCWAKKLNKVSDDKPEYWDTAILETRDCLGATTCDTLGM 120
Db 297 YEEQKGPEVTSNAALTLRNFCWQKQNSPSDRPEHYDTAILTRDCLGSHCTDLGM 356

Qy 121 ADVGTCMDPKRSCSVIEDGLPSAFTTAHELGHVFNPHDNKVKCEEVFGKLRANHMSP 180
Db 357 ADVGTCMDPKRSCSVIEDGLQAAFTTAHELGHVFNPHDDAKHCASLNGVTGSHLMAS 416

Qy 181 TLIQIDRANPSWASAAIITFDLSGHDCLLDOPSKPISLPEDLPGASYTLSQOCELAF 240
Db 417 MLSLDSHQWSPSCSYAYMTSFLDNGHGECLMDKRPQPIKLPDLPTGLVDANKQCQFTF 476

Qy 241 GVSGKPCP-YMQYTKLWCTGKAKQWVCOTRHPFWADGTSCGRGKLCCKGACVERHNLN 299
Db 477 GEESKHCPDAASTCTTLWCCTGSGLLVCOTKHFPFWADGTSCGEGKWCVCVKNTDM- 535

Qy 300 KH---RYDGSWAKWDPYPCSRCTCGGVQGLARRCOTNPTPANGGKYCEGVRYKRSCLNLE 356
Db 536 KHFPATPVHGSWGPWPGWDCSRCTCGGVQVYTMRECDNPVFNNGKGYCEGARVYRSCNIE 595

Qy 357 PCPSSASGKSFREBQ 371
Db 596 DCPDN-NGKTFREBQ 609

RESULT 10
US-10-163-316-7
; Sequence 7, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; FILE REFERENCE: Therefor
; CURRENT APPLICATION NUMBER: US/10/163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus

```

```
Db 373 ADVGTCVDSRSCSVIEDDGLQAAFTTAHELGHVFNMFPHDDAKOCASLNGVNDSHMMAS 432
QY 181 TLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPISLPEDLPASGYTLSSQOCELAF 240
Db 433 MSLNLDHSQSPWSPCSAYMITSLDNGHGECLMDKPNQPIQLPGLPGTSDANRQCQTFE 492
QY 241 GVGSKPCP-YMOYCTKLNCTGKAKQOMVQOTRHPFPWADGTCGEGKLCCLKGACVERHNLN 299
Db 493 GEDSKHCPDAASTCTLMCTGSGGVLVCOTKHFPWADGTCGEGKWCINKCKVNTD-R 551
QY 300 KH---RVDGSWAKWDPYGPCSRTCCGGVQLARRQCTNPTPANGGKYCEGVVRYKRSNCLE 356
Db 552 KHFDTPFHSGWGMGPGWGDSCRTCCGGVQYTMRECDNVPVKNKGKYGKVRVRYKRSNCLE 611
QY 357 PCPSSASGKSFPREQ 371
Db 612 DCPDN-NGKTFREQ 625

RESULT 12
US-10-174-590-352
; Sequence 352, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Goddard, Paul J.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-352

Query Match 57.9%; Score 1181.5; DB 9; Length 837;
Best Local Similarity 55.5%; Pred. No. 1.2e-94;
Matches 208; Conservative 56; Mismatches 106; Indels 5; Gaps 4;

QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPHSILNPINIVVVKVLL 60
Db 213 FASLSRFVETLVVADDKMAAFHAGLKYLLTVMAAAKAFKHSIRNPVSLVTVRLVIL 272
QY 61 RDRDGPVKYTGNAALTLRNFCWQKLNKVS DKHPEYWDTAITLFTRODLGATTCDTLGM 120
Db 273 GSGEGPQVGPAAQTLRSFCAWQRLNTPEDSGPDHFDTAITLFTRODLGCVSTCDTLGM 332
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMFPHDDAKOCASLNGVNDSHMMAS 179
Db 333 ADVGTVCDPARSAIVEDDGLQSAFTAHELGHVFNMLHDSKPCISLNGPLSTSRHVMA 392
QY 180 PTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPISLPEDLPASGYTLSSQOCELA 239
Db 393 PYNAHVDPPEPWSPCSARFIDFLDNGYGHCLLDKPEAPLHPVTFPGKYDADROCOLT 452
QY 240 FGVGSKPCPYM-QYCTKLNCTGKAKQOMVQOTRHPFPWADGTCGEGKLCCLKGACVERHN 298
Db 453 FGPDNRHCPQLPPPCAAALWCSGHLNGHAMCQTKHSPWADGTPCGPAQACMGRCRLHMDQL 512
QY 299 NKHRVD--GSAKWDPYGPCSRTCCGGVQLARRQCTNPTPANGGKYCEGVVRYKRSNCLE 356
Db 513 QDFNTPQAGWGPWGDSCRTCCGGVQVFSRDRCTRPVPPNGGKYCEGRTRFRSCNTE 572
QY 357 PCPSSASGKSFPREQ 371
Db 573 DCP-TCGSAITFPREQ 586

RESULT 14
US-10-175-737-352
```

```
QY 299 NKHRVD--GSAKWDPYGPCSRTCCGGVQLARRQCTNPTPANGGKYCEGVVRYKRSNCLE 356
Db 513 QDFNTPQAGWGPWGDSCRTCCGGVQVFSRDRCTRPVPPNGGKYCEGRTRFRSCNTE 572
QY 357 PCPSSASGKSFPREQ 371
Db 573 DCP-TCGSAITFPREQ 586

RESULT 13
US-10-176-758-352
; Sequence 352, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT FILING DATE: 2002-06-21
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-352

Query Match 57.9%; Score 1181.5; DB 9; Length 837;
Best Local Similarity 55.5%; Pred. No. 1.2e-94;
Matches 208; Conservative 56; Mismatches 106; Indels 5; Gaps 4;

QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPHSILNPINIVVVKVLL 60
Db 213 FASLSRFVETLVVADDKMAAFHAGLKYLLTVMAAAKAFKHSIRNPVSLVTVRLVIL 272
QY 61 RDRDGPVKYTGNAALTLRNFCWQKLNKVS DKHPEYWDTAITLFTRODLGATTCDTLGM 120
Db 273 GSGEGPQVGPAAQTLRSFCAWQRLNTPEDSGPDHFDTAITLFTRODLGCVSTCDTLGM 332
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMFPHDDAKOCASLNGVNDSHMMAS 179
Db 333 ADVGTVCDPARSAIVEDDGLQSAFTAHELGHVFNMLHDSKPCISLNGPLSTSRHVMA 392
QY 180 PTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPISLPEDLPASGYTLSSQOCELA 239
Db 393 PYNAHVDPPEPWSPCSARFIDFLDNGYGHCLLDKPEAPLHPVTFPGKYDADROCOLT 452
QY 240 FGVGSKPCPYM-QYCTKLNCTGKAKQOMVQOTRHPFPWADGTCGEGKLCCLKGACVERHN 298
Db 453 FGPDNRHCPQLPPPCAAALWCSGHLNGHAMCQTKHSPWADGTPCGPAQACMGRCRLHMDQL 512
QY 299 NKHRVD--GSAKWDPYGPCSRTCCGGVQLARRQCTNPTPANGGKYCEGVVRYKRSNCLE 356
Db 513 QDFNTPQAGWGPWGDSCRTCCGGVQVFSRDRCTRPVPPNGGKYCEGRTRFRSCNTE 572
QY 357 PCPSSASGKSFPREQ 371
Db 573 DCP-TCGSAITFPREQ 586

RESULT 14
US-10-175-737-352
```



```
; Sequence 352, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10175.737
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-352

Query Match 57.9%; Score 1181.5; DB 9; Length 837;
Best Local Similarity 55.5%; Pred. No. 1.2e-94;
Matches 208; Conservative 56; Mismatches 106; Indels 5; Gaps 4;

QY 1 FVSIPIRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPHSILNINIVVVKVLL 60
Db 213 FASLSRFVETLVVADDKAAAFHAGLKRKYLITVMAAAKAFKHSIRNPVSLVTVRLVIL 272
QY 61 RDRDSGPKVTGNAALTILNFCAWQKLNKVSCKHPEYWDTAITLFTRODLCGATTCDTLGM 120
Db 273 GSGEEGPQVGPSSAQTLSFCAWQRLNTPEDSGPDHFDTAITLFTRODLCGVSTCDTLGM 332
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVEFGKL-RANHMMS 179
Db 333 ADVGTVCDPARSCAIVEDDGLQSAFTAHELGHVFNHMLHDSKPCISLNGPLSTSRHVMA 392
QY 180 PTLIQIDRANPWSACSAIITDFLDSGHGDCILDDQSPKPSILPEDLPQASYYTLSSQOCELA 239
Db 393 PVMAHVDPEEPWSPCSARFITDFLNGYGHCLLDKPEAPLHPVTFPGKDYDADRQCLT 452
QY 240 FGVGSKPCPYM-QYCTKLWCTGKAKGQMVCTRHFPWADGTSCEGKLCCLKGACVERHNL 298
Db 453 FGPDSRHCPQLPPPCAALWCSGHLNGHAMCQTKHSPWADGTCGPAQACMGGRCLHMDQL 512
QY 299 NKHRVD--GSWAKWDPYPCSRCTCGGVQLARRQCTNPTPANGKYCEGVVRYKRSNLE 356
Db 513 QDFNIPOAGGWGPWGPWGDSCRTCGGVQVFSRDCTRPVPRNGGKYCEGRRTFRSCNTE 572
QY 357 PCPSSASGKSFRREQ 371
Db 573 DCP-TGSALTFRREQ 586

RESULT 15
US-10-173-706-352
; Sequence 352, Application US/10173706
; Publication No. US20030022293A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
```

```
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C7
; CURRENT APPLICATION NUMBER: US/10/173,706
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-706-352

Query Match 57.9%; Score 1181.5; DB 9; Length 837;
Best Local Similarity 55.5%; Pred. No. 1.2e-94;
Matches 208; Conservative 56; Mismatches 106; Indels 5; Gaps 4;

QY 1 FVSIPIRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPHSILNINIVVVKVLL 60
Db 213 FASLSRFVETLVVADDKAAAFHAGLKRKYLITVMAAAKAFKHSIRNPVSLVTVRLVIL 272
QY 61 RDRDSGPKVTGNAALTILNFCAWQKLNKVSCKHPEYWDTAITLFTRODLCGATTCDTLGM 120
Db 273 GSGEEGPQVGPSSAQTLSFCAWQRLNTPEDSGPDHFDTAITLFTRODLCGVSTCDTLGM 332
QY 121 ADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVEFGKL-RANHMMS 179
Db 333 ADVGTVCDPARSCAIVEDDGLQSAFTAHELGHVFNHMLHDSKPCISLNGPLSTSRHVMA 392
QY 180 PTLIQIDRANPWSACSAIITDFLDSGHGDCILDDQSPKPSILPEDLPQASYYTLSSQOCELA 239
Db 393 PVMAHVDPEEPWSPCSARFITDFLNGYGHCLLDKPEAPLHPVTFPGKDYDADRQCLT 452
QY 240 FGVGSKPCPYM-QYCTKLWCTGKAKGQMVCTRHFPWADGTSCEGKLCCLKGACVERHNL 298
Db 453 FGPDSRHCPQLPPPCAALWCSGHLNGHAMCQTKHSPWADGTCGPAQACMGGRCLHMDQL 512
QY 299 NKHRVD--GSWAKWDPYPCSRCTCGGVQLARRQCTNPTPANGKYCEGVVRYKRSNLE 356
Db 513 QDFNIPOAGGWGPWGPWGDSCRTCGGVQVFSRDCTRPVPRNGGKYCEGRRTFRSCNTE 572
QY 357 PCPSSASGKSFRREQ 371
Db 573 DCP-TGSALTFRREQ 586

Search completed: April 29, 2003, 17:21:49
Job time : 12.2469 secs
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 8.58125 seconds
(without alignments)
4156.253 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_583

Perfect score: 2042

Sequence: 1 FVSIPIRYVETLVVADESVMVK.....SCNLEPCPSASGKSPREQ 371

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1337	65.5	951	2 T00017	gene ADAMTS-1 prot
2	1182.5	57.9	837	2 T00355	hypothetical prote
3	923.5	45.2	2165	2 T21371	hypothetical prote
4	684	33.5	550	2 T47158	hypothetical prote
5	628.5	30.8	1205	2 T18517	procollagen N-endo
6	393	19.2	860	2 T16892	hypothetical prote
7	377.5	18.5	1444	2 T18856	angiogenesis inhib
8	297.5	14.6	957	2 T15976	hypothetical prote
9	233.5	11.4	903	2 S60257	meltrin alpha - mo
10	231.5	11.3	508	2 T22836	hypothetical prote
11	229.5	11.2	1170	1 TSHUP1	thrombospondin 1 p
12	224.5	11.0	1170	2 A40558	thrombospondin 1 p
13	220	10.8	571	2 S24789	thrombospondin 1 p
14	218.5	10.7	419	2 S47607	thrombospondin 1 p
15	217.5	10.7	419	2 A59414	thrombospondin 1 p
16	213	10.4	549	2 A8169	thrombospondin 1 p
17	210	10.3	416	2 A37877	thrombospondin 1 p
18	209.5	10.3	617	2 S48160	thrombospondin 1 p
19	208.5	10.2	609	2 S55270	thrombospondin 1 p
20	201.5	9.9	407	2 S66260	thrombospondin 1 p
21	199	9.7	826	2 A60385	thrombospondin 1 p
22	197.5	9.7	411	1 HYSNFA	thrombospondin 1 p
23	197	9.6	480	1 A30065	thrombospondin 1 p
24	196.5	9.6	814	2 G02390	thrombospondin 1 p
25	195.5	9.6	203	2 A59421	thrombospondin 1 p
26	194	9.5	1074	2 JCS928	thrombospondin 1 p
27	193.5	9.5	481	2 JCS432	thrombospondin 1 p
28	191.5	9.4	202	2 JC2350	thrombospondin 1 p
29	191	9.4	478	2 JC4880	thrombospondin 1 p

30 189 9.3 429 2 A42972 coagulation factor
31 187.5 9.2 414 2 S41609 atrolysin C (EC 3.
32 184.5 9.0 481 2 S43125 trimucin precursor
33 181.5 8.9 414 1 HYRSAC atrolysin C (EC 3.
34 181 8.9 600 2 I49281 fertilin alpha pre
35 180.5 8.8 952 2 T18900 disintegrin and me
36 180 8.8 1172 2 A42587 thrombospondin 2 p
37 179.5 8.7 200 2 S15111 hemorrhagic factor
38 177 8.7 1572 2 T00027 brain-specific ang
39 174.5 8.5 1178 1 A39804 thrombospondin pre
40 173 8.5 478 2 A43296 atrolysin E (EC 3.
41 170.5 8.3 469 1 S29126 thrombospondin 2 p
42 170.5 8.3 1172 1 TSHUP2 thrombospondin 2 p
43 170 8.3 202 1 HYRSR ruberlysin (EC 3.4
44 170 8.3 437 2 S05478 properdin - mouse
45 169.5 8.3 1584 2 T00026 brain-specific ang

ALIGNMENTS

RESULT 1

T00017

gene ADAMTS-1 protein - mouse

C:Species: Mus musculus (house mouse)

C>Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000

C:Accession: T00017

R:Kuno, K.; Lizasa, H.; Ohno, S.; Matsushima, K.

Genomics 46, 466-471, 1997

A:Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1

A:Reference number: Z14055; MUID:98110583; PMID:9441751

A:Accession: T00017

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-951 <KUN>

A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BAA24501.1; PID:g2809057

A:Experimental source: strain 129SVJ

C:Genetics:

A:Gene: ADAMTS-1

A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 602/1; 660/3; 719/2

C:Superfamily: thrombospondin type 1 repeat homology

F:542-598/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 65.5%; Score 1337; DB 2; Length 951;
Best Local Similarity 63.2%; Pred No. 3.9e-97;
Matches 237; Conservative 51; Mismatches 81; Indels 6; Gaps 4;
QY 1 FVSIPIRYVETLVVADESVMVKFHGADLEHYLLTFLATAARLYRHPHSILNPINIVVVKVLL 60
Db 237 FVSSPRYVETLVVADESVMVKFHGADLEHYLLTFLATAARLYRHPHSILNPINIVVVKVLL 296
QY 61 RDRSGPKVTGNAALTLRNFCAWOKKLNKVDKHPHYWDTAILFTRODLCGATTCDDTGM 120
Db 297 YEEGKPEYTSNAALTLRNFCAWOKKLNKVDKHPHYWDTAILFTRODLCGATTCDDTGM 356
QY 121 ADVTGMDPCRSKSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEYFGKLRANHMSP 180
Db 357 ADVTGMDPCRSKSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEYFGKLRANHMSP 416
QY 181 TLTIQIDRANPWSACSAIITDFDSHGGLDLPQSPISLPLDLPASGATYLSQCELA 240
Db 417 MLSSLDHSPWSPCSAYMVVTSFLDNGHGBCLMDKPNKIKPLSDLPGLTDANRQCQTF 476
QY 241 GYSGRCP- YMOYCYFKLCTGKAKQMVQTFHFWDGTSGGEGKCLKKGACVRRNLN 299
Db 477 GEESHKCPDAASTCTTLTCTGSGGLVCQTKHFPWADGTSGGEGKCLKKGACVRRNLN 535
QY 300 KH- --RVDSWAKWDPYGPSCRTCGGVLQARRQCTNTPANGGKYCBGVVRYSCNLE 356
Db 536 KHFAFPVHGSGWPGWGDSCSTCGGVQYTMRECDNPVPRNGGKYCBGVVRYSCNLE 595
QY 357 PCPSSASGKSPREQ 371
|| : :||:|||||

Db 596 DCPDN-NGKTFREEQ 609

RESULT 2

T00355

hypothetical protein KIAA0688 - human

C:Species: Homo sapiens (man)

C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000

C:Accession: T00355

R:Shikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.

DNA Res. 5, 169-176, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete

A:Reference number: Z14142; MUID:98403880; PMID:9734811

A:Accession: T00355

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-837 <ISH>

A:Cross-references: EMBL:AB014588; MID:g3327189; PIDN:BAA31663.1; PID:g3327190

A:Experimental source: brain

C:Genetics:

A:Gene: KIAA0688

C:Superfamily: thrombospondin type 1 repeat homology

F:519-575/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 57.9%; Score 1182.5; DB 2; Length 837;

Best Local Similarity 55.5%; Pred. No. 4.7e-85;

Matches 208; Conservative 56; Mismatches 106; Indels 5; Gaps 4;

QY 1 FVSTPRVETLVVADESMVFKHGDLEHYLLTLTAARLYRHPSILNINIVVKKVLL 60

Db 213 FASLSREVELVWADDKAAAFHAGLKRLLTVMAAAKAFKPSIRNPVSLVTVLVL 272

QY 61 RDRSGPKVTGNAALTIRNFCAMQKLNKVSDEKHPYWDTAIFTRDLCGATTCDTGLM 120

Db 273 GSGEGGFQVPSAAQILURSCAMQGLNTPEDSDPDHFDTAIFTRDLCGVTCDTGLM 332

QY 121 ADVGTCMDKRRSCSVIEDDGLPSAFTTAHELGHVFNPHDNNVVCVEVFGKL-RANHMS 179

Db 333 ADVGTCMDPARSCAIVEDDGLQSAFTAAHELGHVFNHNLNHSKPCISLNGPLSTRVMA 392

QY 180 PTLTQIDRANPWSACSAIITDFLDLSDGHCGLDQPSKPSLPEDLPGASYTLSSQCELA 239

Db 393 PVMAHVDPPEPWPSCSARFITDFLDNGYGHCLLDKPEAPLHLPTFPKDYDADROCOLT 452

QY 240 FGVGSKPCPYM-QYCTKLWCTGKAKGMVQCTRHFPWADTSGEGKCLKLGACVERHNL 298

Db 453 FGPDSRHCPQLPPCAALWCSGHLNGHAMCOTKHSPPWADGTPCGPAQACMGRCGLHMDQL 512

QY 299 NKHRVD--GSWAKWDYPGCSRTCGGGVQLARQCTNPTPANGKYCEGVVRYRSCNLE 356

Db 513 QDFNIPQAGGMPGMPGDCSRTCGGGVQFSSRDCTRPVPRNGGKYCEGRRTFRSCNTE 572

QY 357 PCPSASGSKSFREEQ 371

Db 573 DCP-TGSALTTFREEQ 586

RESULT 3

T21371

hypothetical protein F25H8.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T21371; T24896

R:Gajadaty, S.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19413

A:Accession: T21371

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2165 <NLS>

A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3

A:Experimental source: clone F25H8

R:Gajadaty, S.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19949

A:Accession: T24896

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2165 <MI2>

A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3

A:Experimental source: clone T13H10

C:Genetics:

A:Gene: CESP:F25H8.3

A:Map position: 4

A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1;

Query Match 45.2%; Score 923.5; DB 2; Length 2165;

Best Local Similarity 45.4%; Pred. No. 2.9e-64;

Matches 177; Conservative 60; Mismatches 126; Indels 27; Gaps 7;

QY 7 YVETLVVADESMVFKHGDLEHYLLTLTAARLYRHPSILNINIVVKKVLLLRDRDSG 66

Db 281 YVEVLVWADTKMYEYHGRSLEDYVLTFTVASIYRHQSLRASINVVVVKLVLTENAG 340

QY 67 PKVTGNAALTIRNFCAMQKLNKVSDEKHPYWDTAIFTRDLCGAT--TCDTLGMADVGT 125

Db 341 PRITQNAQTLQDFCRWQYYNDPDDSSVQHHDVAILLTRKDICESQKCDTLGLAELGT 400

QY 126 MCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNNVVCVEV-----FGKL 172

Db 401 MCDMQKSCAIIEDNGLSAAFTTAHELGHVFSIPHDERKCYMPVNVKVFQSTKFDKT 460

QY 173 RAN---HMSPPLTQIDRANPWSACSAIITDFLDLSDGHC--DCLLDQPSKPSLPE---- 223

Db 461 QPQNFMHMAPTLEYNTHPWSWSPCSAGMLERFLENNGQTCQLFDQPVERRYEDVFR 520

QY 224 DLPGASYTLSSQCELAFGVSGKPCPYMOCYCTKLWCTGKAKGMVQCTRHFPWADTSGE 283

Db 521 DFPGKKYDAHQCKVFPGPASELCYMPCTCRWLWCATFYGSMGCTQHMPWADGTPCDE 580

QY 284 GK--LCLKGACVERHNLNKHVRDVGSWAKWDYPGCSRTCGGGVQLARQCTNPTPANGGK 341

Db 581 SRSMFCHGACVRLAPESTLKIDGQGDWRWSWEGCSRTCGGVQKGLRDCDSPKPRNGK 640

QY 342 YCEGVVRYRSCNLEPCPSASGSKSFREEQ 371

Db 641 YCVQGRYRSCNTQECQWDT--QYREVQ 668

RESULT 4

T47158

hypothetical protein DKFZp762C1110.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47158

R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.

submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24379

A:Accession: T47158

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-550 <AAA>

A:Cross-references: EMBL:AL162080

A:Experimental source: adult melanoma (Mewo cell line); clone DKFZp762C1110

C:Genetics:

A:Note: DKFZp762C1110.1

Query Match 33.5%; Score 684; DB 2; Length 550;

Best Local Similarity 59.2%; Pred. No. 4.4e-46;

Matches 119; Conservative 25; Mismatches 51; Indels 6; Gaps 4;

QY 175 NHMSPPLTQIDRANPWSACSAIITDFLDLSDGHCGLDQPSKPSLPEDLPGASYTLSSQ 234

Db 10 SHMASMLNLDHDSQWPCSAIYMITSFLDNGHGECLMDKPNQPIQLPGDLPGTSYDANR 69

QY 235 QCELAFGVSGKPCPYMOCYCTKLWCTGKAKGMVQCTRHFPWADTSGEGKCLKLGACV 293

C;Accession: T16892
R;Bentley, D.
submitted to the EMBL Data Library, December 1995
A;description: The sequence of *C. elegans* cosmid T19D2.
A;Reference number: Z18599
A;Accession: T16892
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-860 <R>
A;Cross-references: EMBL:U42846; NID:g1125809; PID:g1125810; PIDN:AAA83600.1; CESP:T16892
C;Genetics:
A;Gene: CESP:T19D2.1
A;Introns: 56/3; 96/1; 171/3; 199/3; 245/2; 293/2; 335/3; 395/3; 426/3; 484/2; 505/3; 536/3; 567/3; 599/3; 629/3; 659/3; 689/3; 719/3; 749/3; 779/3; 809/3; 839/3; 869/3; 899/3; 929/3; 959/3; 989/3; 1019/3; 1049/3; 1079/3; 1109/3; 1139/3; 1169/3; 1199/3; 1229/3; 1259/3; 1289/3; 1319/3; 1349/3; 1379/3; 1409/3; 1439/3; 1469/3; 1499/3; 1529/3; 1559/3; 1589/3; 1619/3; 1649/3; 1679/3; 1709/3; 1739/3; 1769/3; 1799/3; 1829/3; 1859/3; 1889/3; 1919/3; 1949/3; 1979/3; 2009/3; 2039/3; 2069/3; 2099/3; 2129/3; 2159/3; 2189/3; 2219/3; 2249/3; 2279/3; 2309/3; 2339/3; 2369/3; 2399/3; 2429/3; 2459/3; 2489/3; 2519/3; 2549/3; 2579/3; 2609/3; 2639/3; 2669/3; 2699/3; 2729/3; 2759/3; 2789/3; 2819/3; 2849/3; 2879/3; 2909/3; 2939/3; 2969/3; 2999/3; 3029/3; 3059/3; 3089/3; 3119/3; 3149/3; 3179/3; 3209/3; 3239/3; 3269/3; 3299/3; 3329/3; 3359/3; 3389/3; 3419/3; 3449/3; 3479/3; 3509/3; 3539/3; 3569/3; 3599/3; 3629/3; 3659/3; 3689/3; 3719/3; 3749/3; 3779/3; 3809/3; 3839/3; 3869/3; 3899/3; 3929/3; 3959/3; 3989/3; 4019/3; 4049/3; 4079/3; 4109/3; 4139/3; 4169/3; 4199/3; 4229/3; 4259/3; 4289/3; 4319/3; 4349/3; 4379/3; 4409/3; 4439/3; 4469/3; 4499/3; 4529/3; 4559/3; 4589/3; 4619/3; 4649/3; 4679/3; 4709/3; 4739/3; 4769/3; 4799/3; 4829/3; 4859/3; 4889/3; 4919/3; 4949/3; 4979/3; 5009/3; 5039/3; 5069/3; 5099/3; 5129/3; 5159/3; 5189/3; 5219/3; 5249/3; 5279/3; 5309/3; 5339/3; 5369/3; 5399/3; 5429/3; 5459/3; 5489/3; 5519/3; 5549/3; 5579/3; 5609/3; 5639/3; 5669/3; 5699/3; 5729/3; 5759/3; 5789/3; 5819/3; 5849/3; 5879/3; 5909/3; 5939/3; 5969/3; 5999/3; 6029/3; 6059/3; 6089/3; 6119/3; 6149/3; 6179/3; 6209/3; 6239/3; 6269/3; 6299/3; 6329/3; 6359/3; 6389/3; 6419/3; 6449/3; 6479/3; 6509/3; 6539/3; 6569/3; 6599/3; 6629/3; 6659/3; 6689/3; 6719/3; 6749/3; 6779/3; 6809/3; 6839/3; 6869/3; 6899/3; 6929/3; 6959/3; 6989/3; 7019/3; 7049/3; 7079/3; 7109/3; 7139/3; 7169/3; 7199/3; 7229/3; 7259/3; 7289/3; 7319/3; 7349/3; 7379/3; 7409/3; 7439/3; 7469/3; 7499/3; 7529/3; 7559/3; 7589/3; 7619/3; 7649/3; 7679/3; 7709/3; 7739/3; 7769/3; 7799/3; 7829/3; 7859/3; 7889/3; 7919/3; 7949/3; 7979/3; 8009/3; 8039/3; 8069/3; 8099/3; 8129/3; 8159/3; 8189/3; 8219/3; 8249/3; 8279/3; 8309/3; 8339/3; 8369/3; 8399/3; 8429/3; 8459/3; 8489/3; 8519/3; 8549/3; 8579/3; 8609/3; 8639/3; 8669/3; 8699/3; 8729/3; 8759/3; 8789/3; 8819/3; 8849/3; 8879/3; 8909/3; 8939/3; 8969/3; 8999/3; 9029/3; 9059/3; 9089/3; 9119/3; 9149/3; 9179/3; 9209/3; 9239/3; 9269/3; 9299/3; 9329/3; 9359/3; 9389/3; 9419/3; 9449/3; 9479/3; 9509/3; 9539/3; 9569/3; 9599/3; 9629/3; 9659/3; 9689/3; 9719/3; 9749/3; 9779/3; 9809/3; 9839/3; 9869/3; 9899/3; 9929/3; 9959/3; 9989/3; 10009/3; 10039/3; 10069/3; 10099/3; 10129/3; 10159/3; 10189/3; 10219/3; 10249/3; 10279/3; 10309/3; 10339/3; 10369/3; 10399/3; 10429/3; 10459/3; 10489/3; 10519/3; 10549/3; 10579/3; 10609/3; 10639/3; 10669/3; 10699/3; 10729/3; 10759/3; 10789/3; 10819/3; 10849/3; 10879/3; 10909/3; 10939/3; 10969/3; 10999/3; 11029/3; 11059/3; 11089/3; 11119/3; 11149/3; 11179/3; 11209/3; 11239/3; 11269/3; 11299/3; 11329/3; 11359/3; 11389/3; 11419/3; 11449/3; 11479/3; 11509/3; 11539/3; 11569/3; 11599/3; 11629/3; 11659/3; 11689/3; 11719/3; 11749/3; 11779/3; 11809/3; 11839/3; 11869/3; 11899/3; 11929/3; 11959/3; 11989/3; 12019/3; 12049/3; 12079/3; 12109/3; 12139/3; 12169/3; 12199/3; 12229/3; 12259/3; 12289/3; 12319/3; 12349/3; 12379/3; 12409/3; 12439/3; 12469/3; 12499/3; 12529/3; 12559/3; 12589/3; 12619/3; 12649/3; 12679/3; 12709/3; 12739/3; 12769/3; 12799/3; 12829/3; 12859/3; 12889/3; 12919/3; 12949/3; 12979/3; 13009/3; 13039/3; 13069/3; 13099/3; 13129/3; 13159/3; 13189/3; 13219/3; 13249/3; 13279/3; 13309/3; 13339/3; 13369/3; 13399/3; 13429/3; 13459/3; 13489/3; 13519/3; 13549/3; 13579/3; 13609/3; 13639/3; 13669/3; 13699/3; 13729/3; 13759/3; 13789/3; 13819/3; 13849/3; 13879/3; 13909/3; 13939/3; 13969/3; 13999/3; 14029/3; 14059/3; 14089/3; 14119/3; 14149/3; 14179/3; 14209/3; 14239/3; 1426

A:Gene: CESP:119D2.1
A: Introns: 56/3: 96/

```

Query Match      19.2%; Score 393; DB 2; Length 860;
Best Local Similarity 29.88; Pred. No. 5.2e-23;
Matches 117; Conservative 54; Mismatches 161; Indels 60; Gaps 19;

Qy      8 VETLVVADESM---VFPHGADLEHYLLTLTLATAAR---LYRHPSILNPINIVVVKVLL 59
      ||| ||| : : : : : ||| : : : ||| : : :
Db      135 VELAVFADDAWMDHFKNMYGKAAEENMHTFTMAVVNNIDVLYTQRLQPRINIKIVREYI 194

Qy      60 LRD----RDSPKVTGNAALTLRNFCQWOKLKNVSKDHPYWDTAILFTQDLCGATTC 115
      ||| : : : ||| : : : ||| : : : ||| : : :
Db      195 LKNTPLHLMNARKHNGVDVRLDLDAFCQYQNEPNPNADPRHWDHALLESYDILHRNGVK 254

Qy      116 DTGLGMADVTGTCMDPKRSCVIEDDGLPSAFTHAELGHVFNNPHNDNVKVCSEVFCKLRAN 175
      ||| ||| : : : ||| : : : ||| : : : |||
Db      255 TVACYAPVKMGCSVRSCITINEGLDFGVSFVVTTHMGHSLGMYHDGDNEDC-----LRC- 308

Qy      176 HMMGPTLIQDRANPWVSACSAAITDFLDSCG-GD-----CLLD---QPSKPISLP-D 224
      :||| : : : ||| : : : ||| : : : ||| : : :
Db      309 CIMGFSV--GSGKTHWSQCSYNEMATEFV--GHLGDFFRPPNCLQDASANEQRMVAFKESE 364

```

QY 225 LPGASYITQQCEBLAFGWGRPC-----PIMQICITATWCGNAGQWVCQIRHFPW 273

D6 365 PPQGLFTLDEQCEIFHG-----ECWKHELKDGQTMONICQMVWC-----GNGEVIRTAH-PA 416

[illegible]

RESULT 7

T18856
angiogenesis inhibitor homolog - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*
C.date: 15-Oct-1999 #sequence revisions 15-Oct-1999 #cont changes 19-Feb-2000

C;Accession: T18856; T24653

submitted to the EMBL Data Library, July 1995

A;Reference Number: Z19031
A;Accession: T18856

A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA

A;Residues: 1-1444 <WIL>

A;Experimental source: clone C02B4

R. McMurry, A.
submitted to the EMBL Data Library, July 1995

A;reference number: Z1991/
A:Accession: T24653

A; Status: preliminary; translated from GB/EMBL/DBJ
A; Molecule type: DNA

A;Residues: 1-1444 <WI2>

A; Experimental source: clone T07C5

C; GENETICS:
A; Gene: CESP:C02B4.1

A:Map position: X
A:Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3; 568

Query Match 18.5%; Score 377.5; DB 2; Length 1444;
Best Local Similarity 30.5%; Pred. No. 1.9e-21;
Matches 115; Conservative 45; Mismatches 170; Indels 47; Gaps 15;

QY 25 DLEHYL---LTLATAARLYRHPSILNPIVIVVKKVLLLRDRS-----GPKVTGNAALT 76
DB 239 DAEQHLLEFSLALNNVHVLYQQDTLPNLDIVIVRYEMWRTQPSALSTGVHKNQQAQSL 298
QY 77 LRNFCAWQKLNKYSV---KHPEYWDATLFTQDLCCGATTCDTLGMADVGMCDPKRSCSV 135
DB 299 LDAFCRYQAHNPNPGLTDMNHYDGVLLTGYDIYHTT---SVAGVAPVARMCDPLFACSL 357
QY 136 IEDDGLPSAFTAHLELGHVFNPHDNV---VCEEVFGKLRANHMSPPLIQIDRANPHSAC 194
DB 358 VEGHLGRSFVLAHEMGMHGMVHDGVQNCCKCLMSAYNGAGKT-----TWSDC 409
QY 195 SAAIITDFL---DSGHGDCILDPQSKPISLPE-----DLPGASYTLQQCELAFAFGVSK- 245
DB 410 SVREFNAFLQLDDESGRNCILRDSPLGISTNHLSDLPLPQORTAQOCSYFWGROYKV 469
QY 246 PCP-----YMOYCTKLWCTKAKGQMVQCOTRHPFWDGTSCEGKCLKGACVE---RHNLNK 300
DB 470 EIPNGKAMDDICRILWCGNSGS---TISTAH-PALEGSWCGANKWCHKGCQCTHTWTFGLTP 525
QY 301 HRVDSWAKWD-----PYGCSRTCAGGVOLARQCTNPTPANGGKYCEGVRYKYRSCN 354
DB 526 VPIDGSEWGAEGKCPICQCAVSGSITVQGHRCVNPAPNNGGKTCEGANIRGIVCG 585
QY 355 LEPCPSSASGKSFREQ 371
DB 586 ----ATSSNCLGFTREE 598

RESULT 8
Ti5976
hypothetical protein F08C6.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: Ti5976
R:Bentley, D.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid F08C6.
A:Reference number: 218440
A:Accession: Ti5976
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-957 <GEN>
A:Cross-references: EMBL:U29378; NID:g868184; PID:g868185; PIDN:AAA68721.1; CESP:F08C6.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP-F08C6.1
A:Introns: 23/1; 135/3; 220/3; 285/3; 325/2; 376/3; 461/2; 498/3; 532/3; 627/2; 714/1; 7

Query Match 14.6%; Score 297.5; DB 2; Length 957;
Best Local Similarity 25.2%; Pred. No. 1.9e-15;
Matches 116; Conservative 54; Mismatches 157; Indels 133; Gaps 19;

QY 8 VETLVVADSWKVF---HG--AD--LEHYLLTLATAARLYRHPSILNPIVIVVKKVLL 59
DB 221 VELAVFVDENLWRFHSSKGMADRKLQDTLTLLNNIQIYQTPASPLFRVIREV 280
QY 60 LRDRDSG---PKVTGNAALTLRNFCAWQKLNKYSVKHPEYWDATLFTQDL-CCGATT 114
DB 281 LTRQPSALACYLHNHNGNAQWYLDRCRYQNL-VRD-----WDHAIMLTGYDIHRGAGS 334
QY 115 CDTLGMADVGMCDPKRSCSVIEDDGLPSAFTAHLELGHVFNPHDNVKNVCEVFGKLA 174
DB 335 RSISGIARLDGMCDPNTCTLAEGDLSAFTIGHELGRHFTLKD-----KS 383
QY 175 NHMMSPTL-----IQIDRANPHSACS-----AAIIT-----DFLDSGHGDCILD 213

DB 384 DTLTFTGFCGSKWQGLGRVCPWTGTNETIQTQVHVAPVVTTLPSRIDGSMGSGWATICS 443
QY 214 Q-----PSKPISLPEDLPAGS-----YTLSQOCELA----- 239
DB 444 QCTCNGILSGVGLAIARRTCSAPYPANGGSDCVGSTSRVLCRQGRASKSVDEYISDK 503
QY 240 -----FGVSGKPCPYMOYCTKLWCTKAKGQMVQCOTRHPF-----ADGTSC 281
DB 504 CMEQKRLKNDRELTKGGSQNLNRPQARACKVFCVDQ---QHYGSQRNRYRFGDNLDPGTSC 560
QY 282 GEGKCLKGACVHRHNLKHRVD-----GSAWAKWDYPGCSRTC 320
DB 561 GYDEYCLDGECLALNCNNALISRDQSCPTDCTPITDQSSSVYRGWGTWSLWTSCTATC 620
QY 321 GGGVOLARQCTNPTPANGGKYCEGVRYKYRSCNLEPCPS 360
DB 621 GGGYRKNRACS-----ITG--QCEGNEDETEVCSSSESCPS 654

RESULT 9
S60257
meltrin alpha - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 20-Jun-2000
R:Yagami-Hiromasa, T.; Sato, T.; Kurisaki, T.; Kamijo, K.; Nabeshima, Y.I.; Fujisawa-Nature 377, 652-656, 1995
A:Title: A metalloprotease-disintegrin participating in myoblast fusion.
A:Reference number: S60257; MUID:96026308; PMID:7566181
A:Accession: S60257
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-903 <YAG>
A:Cross-references: EMBL:D50411; NID:g1054586; PIDN:BAA08912.1; PID:g1054587
C:Superfamily: mouse meltrin alpha; disintegrin homology
F:421-503/Domain: disintegrin homology <Dis>
F:349/Active site: Glu #status predicted

Query Match 11.4%; Score 233.5; DB 2; Length 903;
Best Local Similarity 21.1%; Pred. No. 1.9e-10;
Matches 107; Conservative 49; Mismatches 149; Indels 203; Gaps 16;

QY 2 VSIPIRYVETLVVADSWKVFHGDLEHYLLTLATAARLYRHPSILNPIVIVVKKVLLR 61
DB 208 LKMTKYVELVIVADNREFQRGDLEKVKQRLIEIANHVDKFFRPLN-IRIVLGVGVWN 266
QY 62 DRDSGPKVTGNAALTLRNFCAWQKLNKYSVKHPEYWDATLFTQDLCCGATTCDTLGM 121
DB 267 DIDR-CSIQDPFTRLHEFLDWRKIKLLPRKSH-----DNAQLISGVYFQGT----TIGMA 317
QY 122 DVTMCDPKRSCSVI---EDDGLPSAFTAHLELGHVFNPHDNV-KVCEEVFGKLRANHM 177
DB 318 PIMSMCTAEQSGVYMDHSDSPLAGAAYTLAHELGHNFEGMNHDTLERCSCSRMAEKGGCI 377
QY 178 MSPTLIQIDRANP---WSACSAIIITDFLDSGHGDCILDPQ----- 215
DB 378 MNPS-----TCFPPPMVFSSCRKDLKLEASLEKGMGMCLEFNLPEVKQAFGKRGKGVYEE 432
QY 216 -----SKPISLPED 224
DB 433 GEECDGCEPECTNRCCNATTCTLKPDVCAHGQCCEDCOLKPGGTACRGSSNSCDLPEF 492
QY 225 LPGAS-----YTLSQOCELAFCGSKGPCPYMOY----- 252
DB 493 CTGTAPECPANVYLHDGHPGQGVGYNGICQTHEQQCVTLGWGPAGKAPGICFERVNS 552
QY 253 -----CTKLWCTGKAK-----GOMV 267
DB 553 AGDPYGNCGKDSKSAFAKCELDAKCGKIQCGGASRPVIGTNAVSIETNIPQOEGGRIL 612
QY 268 CQTRHFPWAD-----GTSCEGKCLKGACVHRHNLKHRVDGSWAKWDYPGPCS 317
DB 317 -----

QY 120 MADVGTMCDFKRSVIEDDGLPSAFTTAHELGHVFNPHDNV-KVCEEVFGKLRANHHM 178
Db 278 -----SSWVLELRLGLTIIVTTL-----QDSIRKVTTEE--NKELANELR 313
QY 179 SPTL-----IQIDRANPWSA-----CSAAIITDFLDSGHGD 209
Db 314 RPPLCYHNGVQYRNEEWTVDSCTECHCONSITCKKVSCTPIMPCSNATVPD-----GE 367
QY 210 CLLDQPSKPSLSPED-----LPGASYT-----LSQOCELAFVGS 244
Db 368 CC-----PRCWPDSADDSGSPWSEWTSCTSCGNGIQGRSCDSLNRCE-GSSVQT 420
QY 245 KPCPYMOYCTKLWCTGKAKGMVQCTRHFPWAD-GTSCGEG----- 284
Db 421 RTC-HIQECDKRF---KQDGGW---SHWSPWSSCVTCGDGVITRIRLNCSPQMNGKP 473
QY 285 -----KLCIKGACVERHNKRVHVDGWSAKWDPYGPCSRTCGGVLARROCTNPTP 336
Db 474 CEGARETKACKKDAC-----PINGGWPSPWDICSVTCGGGVQKRSRLCNPPTP 524
QY 337 ANGKGYCEGVYKVRSCNLEPCP 359
Db 525 QFGGKDCVGDVTENQICNKQDCP 547

RESULT 12
A40558
thrombospondin 1 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 20-Aug-1999
C:Accession: A40558; B437905; B42587; S68787
R:Lawler, J.; Duquette, M.; Ferro, P.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.
Genomics 11, 587-600, 1991
A:Title: Characterization of the murine thrombospondin gene.
A:Reference number: A40558; MUID:92128941; PMID:1774063
A:Accession: A40558
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1170 <LAW>
A:Cross-references: GB:M62449; GB:M62450; GB:M62451; GB:M62452; GB:M62453; GB:M62454; GB:M62455; GB:M62466; GB:M62467; GB:M62468; GB:M62469; GB:M62470; NID:9511867; PIDN:AAA5061
R:Bornstein, P.; Alfli, D.; Devayayalu, S.; Franson, P.; Li, P.
J. Biol. Chem. 265, 16691-16698, 1990
A:Title: Characterization of the mouse thrombospondin gene and evaluation of the role of
A:Reference number: A37905; MUID:90375546; PMID:2398070
A:Accession: A37905
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <BOR>
A:Cross-references: GB:J05605; GB:J05606; NID:9201991; PIDN:AAA40431.1; PID:g554390
R:Laherty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A:Title: Characterization of mouse thrombospondin 2 sequence and expression during cell
A:Reference number: A42587; MUID:92147683; PMID:1371115
A:Accession: B42587
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1152, 'p', 1154-1170 <LAH>
A:Cross-references: GB:H87276
A:Note: sequence extracted from NCBI backbone (NCBIP:81501)
R:Chen, H.; Aeschlimann, D.; Nowlen, J.; Mosher, D.F.
FEBS Lett. 387, 36-41, 1996
A:Title: Expression and initial characterization of recombinant mouse thrombospondin 1
A:Reference number: S68787; MUID:96234006; PMID:8654563
A:Accession: S68787
A:Molecule type: protein
A:Residues: 19-26, 'X', 28-37 <CHE>
C:Complex: homotrimer, disulfide linked
C:Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; vd
C:Keywords: calcium binding; glycoprotein; homotrimer
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-1170/Product: thrombospondin 1 #status predicted <MAT>
F:317-375/Domain: von Willebrand factor type C repeat homology <VWC>

F:378-429/Domain: thrombospondin type 1 repeat homology <THR1>
F:434-490/Domain: thrombospondin type 1 repeat homology <THR2>
F:491-547/Domain: thrombospondin type 1 repeat homology <THR3>
F:551-586/Domain: EGF homology <EGF>
F:248,360,708,1067/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 11.08; Score 224.5; DB 2; Length 1170;
Best Local Similarity 24.3%; Pred. No. 1.3e-09;
Matches 93; Conservative 35; Mismatches 106; Indels 149; Gaps 20;

QY 63 RDSGPKVTGNAALTNFCAMQKLNKVSXKHPEVDTAILFTTRQDL---CGATTCTDTLG 119
Db 228 NKGSSSTNVLITLDN-----NVNGSPAIRNYIGHKTKDLQAICG-LSCDEL- 277
QY 120 MADVGTMCDFKRSVIEDDGLPSAFTTAHELGHVFNPHDNV-KVCEEVFGKLRANHHM 178
Db 278 -----SSWVLELRLGLTIIVTTL-----QDSIRKVTTEE--NRELVELSK 313
QY 179 SPTL-----IQIDRANPWSA-----CSAAIITDFLDSGHGD 209
Db 314 RPPLCYHNGVQYRNEEWTVDSCTECHCONSITCKKVSCTPIMPCSNATVPD-----GE 367
QY 210 CLLDQPSKPSLSPED-----LPGASYT-----LSQOCELAFVGS 244
Db 368 CC-----PRCWPDSADDSGSPWSEWTSCTSCGNGIQGRSCDSLNRCE-GSSVQT 420
QY 245 KPCPYMOYCTKLWCTGKAKGMVQCTRHFPWAD-GTSCGEG----- 284
Db 421 RTC-HIQECDKRF---KQDGGW---SHWSPWSSCVTCGDGVITRIRLNCSPQMNGKP 473
QY 285 -----KLCIKGACVERHNKRVHVDGWSAKWDPYGPCSRTCGGVLARROCTNPTP 336
Db 474 CEGARETKACKKDAC-----PINGGWPSPWDICSVTCGGGVQKRSRLCNPPTP 524
QY 337 ANGKGYCEGVYKVRSCNLEPCP 359
Db 525 QFGGKDCVGDVTENQICNKQDCP 547

RESULT 13
S24789
jararagin C precursor - jararaca (fragment)
N:Alternate names: single chain botrocetin
C:Species: Bothrops jararaca (Jararaca)
C:Date: 20-Feb-1995 #sequence_revision 29-Aug-1997 #text_change 09-Jun-2000
C:Accession: S24789; JC2245; A44463; A37958; JC2373
R:Paine, M.J.I.
submitted to the EMBL Data Library, August 1992
A:Reference number: S24789
A:Accession: S24789
A:Molecule type: mRNA
A:Residues: 1-571 <PAI>
A:Cross-references: EMBL:X68251; NID:g62467; PID:g62468
R:Usami, Y.; Fujimura, Y.; Miura, S.; Shima, H.; Yoshida, E.; Yoshio, K.
Biochem. Biophys. Res. Commun. 201, 331-339, 1994
A:Title: A 28 kDa protein with disintegrin-like structure (jararagin-C) purified fro
A:Reference number: JC2245; MUID:94256999; PMID:8198592
A:Accession: JC2245
A:Molecule type: protein
A:Residues: 360-571 <USA>
A:Experimental source: venom
R:Paine, M.J.; Desmond, H.P.; Theakston, R.D.; Crampton, J.M.
J. Biol. Chem. 267, 22869-22876, 1992
A:Title: Purification, cloning, and molecular characterization of a high molecular we
ily.
A:Reference number: A44463; MUID:93054601; PMID:1385408
A:Accession: A44463
A:Molecule type: mRNA
A:Residues: 1-23, 'Q', 25-92, 'G', 94-131, 'G', 133-169, 'Q', 171-571 <PA>
A:Cross-references: GB:X68251
A:Experimental source: venom gland
A:Note: sequence inconsistent with the nucleotide translation

A>Note: sequence extracted from NCBI backbone (NCBIP:118104)
R:Fujimura, Y.; Titani, K.; Usami, Y.; Suzuki, M.; Oyama, R.; Matsui, T.; Fukui, H.; Sug
Biochemistry 30, 1957-1964, 1991
A:Title: Isolation and chemical characterization of two structurally and functionally di
A:Reference number: A37958; MUID:91129280; PMID:1993206
A:Accession: A37958

A:Molecule type: protein
A:Residues: 360-372, 'P', 374-378, 'X', 380-384 <FU>
A:Note: 361-Val was also found
C:Comment: Inhibits collagen- and ADP-induced platelet aggregation.
C:Superfamily: mouse meltrin alpha; disintegrin homology
F:118-198/Disulfide bonds: #status experimental
F:158-165/Disulfide bonds: (or 158-160) #status experimental
F:181/Modified site: glycosylated (Asn) #status predicted
F:295,299,305/Binding site: zinc (His) #status predicted
F:296/Active site: Glu #status predicted

Query Match 10.8%; Score 220; DB 2; Length 571;
Best Local Similarity 25.8%; Pred. No. 1.3e-09;
Matches 96; Conservative 50; Mismatches 144; Indels 82; Gaps 21;

QY 6 RYVETLVVADESMVKFHGADLEHYLLTLAARLYRHPSTLNPI-NIVVVKVLL--LRD 62
Db 159 KYIEFVVVDGTVTKNGDLD-----KIKARMYELANIVNEIFRYLYMHVALVGL 211
QY 63 RDSGPKVT--GNAALTLRNFCWOKKLKNSDKHPEYWDTAIFLTRODLCGATTCDTLGM 120
Db 212 WNSGDKITVKPDVDTYLSFAEWRKTDLLTRKKH---DNAQLLTAIDFNG----PIGY 263
QY 121 ADVGTMCDPKRSVIED---DGLPSAFTTAHELGHVFNMPHDNVKVCVEEFGKLRANH- 176
Db 264 AYGSMCHPKRSVIGVQDYSPINLVAVIMAHGHNLIHHD-----GSCSGDY 315
QY 177 --MMSPTLIQDRANPWSACSAIITDLDGSHGDCILDDP-----SKPISLPEDLPGA 228
Db 316 PCIMGPT-INSPEKFFSNCYSIQCFWFMHNPECIINEPLGTDIISPPVCGNELL--- 371
QY 229 SYTLSQCELAFGVSGKPCPYMYCTKLWCTGKAGQVMQCTRHFPWADGTSCEGKLC 288
Db 372 --EVGECD-----CGTPENCQNECCDA-----ATCKLK-----SGSQCGHGD--- 407
QY 289 KGACVERHNLKRVG--GSWAKWDPPYGPC---SRTCGGVQLARRQCTNPTTPANGK 343
Db 408 ---CEEQCKFKSTECRASMSSECDPAEHCTGQSSECPADVFKNGQ---PCLDNYG-YC 460
QY 344 --EGVRKYRSC 353
Db 461 YNGMCPIMYHOC 472

RESULT 14
S41607
atrolysin A (EC 3.4.24.1) - western diamondback rattlesnake (fragment)
N:Alternate names: hemorrhagic toxin a
C:Species: Crotalus atrox (western diamondback rattlesnake)
C:Date: 29-Sep-1994 #sequence_revision 13-Mar-1997 #text_change 09-Jun-2000
C:Accession: S41607
R:Hite, L.A.; Jia, L.G.; Bjarnason, J.B.; Fox, J.W.
Arch. Biochem. Biophys. 308, 182-191, 1994
A:Title: cDNA sequences for four snake venom metalloproteinases: structure, classificati
A:Reference number: S41607; MUID:94145078; PMID:8311451
A:Accession: S41607
A:Status: preliminary; translation not shown
A:Molecule type: mRNA
A:Residues: 1-419 <HIT>
A:Cross-references: EMBL:U01234; NID:g402257; PID:g402258
C:Superfamily: mouse meltrin alpha; disintegrin homology
C:Keywords: hydrolase; metalloproteinase
F:203-291/Domain: disintegrin homology <DIS>
F:143/Active site: Glu #status predicted

Query Match 10.7%; Score 218.5; DB 2; Length 419;
Best Local Similarity 23.8%; Pred. No. 1.2e-09;

Matches 103; Conservative 45; Mismatches 108; Indels 177; Gaps 23;
QY 6 RYVETLVVADESM--VKPHG--ADLEHYLLTLAARLYRHPSTLNPI-NIVVVKVLL--- 59
Db 6 RYVELVIVADHRMFTKYNGNLKKIRKIYQIVNTINEIY-----IPLNIRVALVLEIW 59
QY 60 ----LRDRSGPKVTGNAALTLRNFCWQ--KKLNKVSCKHPEYWDTAIFLTRODLCGAT 113
Db 60 SNGDLID-----VTSANVTLSKSGFNWRYTNLLRRKSH-----DNAQLLTAIDL--- 103
QY 114 TCDTGLMADVGTMCDPKRSVIEDD---GLPSAFTTAHELGHVFNMPHDNVKVCVEEFG 170
Db 104 DEETLGLAPLTGTCDPKLSGIVQDHSPIINLVAVTMAHELGNLGVHDENR-C----- 157
QY 171 KLRANHMMSPTLIQ--IDRANP--WSACSAIITDLDGSHGDCILDDP-----SKPI 219
Db 158 ----HGSTPACVMCAVLRQRPSEFSDCSLNHYRTFIINYPQCILNEPLQTDIISPPV 212
QY 220 SLPEDLP-----GASYT-----LSQOCELA----- 239
Db 213 CGNELLEVGEEDCGSPRTCDPCDDAATCKLHSHWECSEGCCQCKFTSAGNVCRPAR 272
QY 240 -----FGVGSKP-----CPYMQY-CTKLIW----- 257
Db 273 SECIDIAESCTGQSADCTDDFHRNGPKLHNFYCYNGNCPIMYHOCYALWGSNVIVAPD 332
QY 258 -----CTGK--AKQMVCQTRHF-----PWAD-----GTSC 281
Db 333 ACFDINOSGNNSFYCRKENGVINPQAEQEDVKGRLEFCNVNDFLCRHKYSDDGMVDHGTCK 392
QY 282 GEGKCLCKGACVE 294
Db 393 ADGKVKCNROCDV 405

RESULT 15

A59414

metalloproteinase (EC 3.4.24.-) (brevilysin) H6, venom - Gloydus halys brevicaudus
N:Contains: disintegrin
C:Species: Gloydus halys brevicaudus
C:Date: 01-Mar-2002 #sequence_revision 01-Mar-2002 #text_change 01-Mar-2002
C:Accession: A59414
R:Terada, S.
J. Biochem. 128, 167-173, 2000
A:Title: Primary Structure and Autoproteolysis of Brevilysin H6 from the Venom of Glo

A:Reference number: A59414

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-419 <TER>

C:Keywords: hydrolase; metalloproteinase; venom; zinc

F:118-198/Disulfide bonds: #status experimental

F:158-165/Disulfide bonds: (or 158-160) #status experimental

F:181/Modified site: glycosylated (Asn) #status predicted

F:182-187/Disulfide bonds: #status experimental

Query Match

10.7%; Score 217.5; DB 2; Length 419;

Best Local Similarity 21.9%; Pred. No. 1.5e-09;

Matches 96; Conservative 43; Mismatches 119; Indels 181; Gaps 16;

QY 6 RYVETLVVADESMVKFHGADLEHYLLTLAARLYRHPSTLNPI-NIVVVKVLL 59

Db 7 RYVELVIVADKGMVTKNGDLD-----KIKRMVELANNDIYRYMYIHVALVGEI 59

QY 60 LRDRSGPKVTGNAALTLRNFCWOKKLKNSDKHPEYWDTAIFLTRODLCGATTCDTIG 119

Db 60 WSDGDK-ITVTNPVDDTLLSSFAEWRKTDLLTRKKH---DNAQLLTAIDFNG----PIG 110

QY 120 MADVGTMCDPKRSVIED---DGLPSAFTTAHELGHVFNMPHD-NVKVCVEEFGKLRAN 175

Db 111 YAIASMCMPKRSVIGVQDYSPINLVSYVMAHGHNLGIHHDGSGYSCGDIACIMGAT 170

Fri May 2 10:41:46 2003

```
QY 176 HMMSPTLIQIDRANPWSAGSAAIITDFLDSHGDCILDQP-----SKPI----- 219
Db 171 ISHEPSTF-----FSNCSYIOCWDFINDHNPFCIVNEPLGTDIVSPVCGNELLEVGE 223
QY 220 -----SLPEDLPASYT 231
Db 224 EDCGTPENCNECCDAATCKLKSGSQGHGDCCEQCKFSKSGTECRESMPECDPAEHT 283
QY 232 -LSQOCEL-AFGVGSKPC-----P 248
Db 284 QSSSECPADVHKNQOPCLDNTGYCYNGNCPIMYHOCYALWGADVYDAEDSCFESNTKGN 343
QY 249 YMOYCTK-----LMCTGKAKGQ-----MYCQTRHFPW 275
Db 344 YGYCRKENGIKIPCAPEDVKCGRLYCKDNSPGQNNPCKMFYSNEDEHKGW----- 395
QY 276 ADGTSCEGKCLKGACVE 294
Db 396 LPGTKGDKGKVCNGHCVD 414
```

Search completed: April 29, 2003, 17:13:21
Job time : 12.5813 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 4.56924 Seconds
(without alignments)
3367.676 Million cell updates/sec

Title: US-10-009-332-1_COPY_213_583

Perfect score: 2042

Sequence: 1 FVSIPIRYVELVVADESWK.....SCNLEPCPSASGSKSFREQ 371

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1337	65.5	968	1 ATSL_MOUSE	P97857 mus musculus
2	1334	65.3	967	1 ATSL_HUMAN	Q9uh18 homo sapien
3	1327	65.0	967	1 ATSL_RAT	Q9wugl rattus norv
4	1182.5	57.9	837	1 ATSL_HUMAN	Q75173 homo sapien
5	1133.5	56.5	630	1 ATSL_RAT	Q9esp7 rattus norv
6	1128	55.2	905	1 ATSL_MOUSE	P57110 mus musculus
7	1126.5	55.2	890	1 ATSL_HUMAN	Q9up79 homo sapien
8	1114	54.6	930	1 ATSL_MOUSE	Q9r001 mus musculus
9	1110	54.4	930	1 ATSL_HUMAN	Q9una0 homo sapien
10	1046.5	51.2	1629	1 ATSL_HUMAN	Q9p2n4 homo sapien
11	819	40.1	997	1 ATSL_HUMAN	Q9ukp4 homo sapien
12	786	38.5	1593	1 ATSL_HUMAN	P58397 homo sapien
13	751	36.8	1077	1 ATSL_HUMAN	Q9h324 homo sapien
14	633.5	31.0	1211	1 ATSL_HUMAN	O95450 h adamts-2
15	631.5	30.9	1205	1 ATSL_HUMAN	O15072 homo sapien
16	628.5	30.8	1205	1 ATSL_BOVIN	P79331 b adamts-2
17	620	30.4	860	1 ATSL_HUMAN	Q9ukp5 homo sapien
18	599	29.3	207	1 ATSL_BOVIN	Q9tt92 bos taurus
19	599	29.3	245	1 ATSL_BOVIN	Q9tt93 bos taurus
20	246.5	12.1	776	1 AD28_MACFA	Q9xl16 macaca fasc
21	236.5	11.6	909	1 AD12_HUMAN	O43184 homo sapien
22	234	11.5	775	1 AD28_HUMAN	Q9ukq2 homo sapien
23	233.5	11.4	903	1 AD12_MOUSE	O61824 mus musculus
24	229.5	11.2	1170	1 TSPI_HUMAN	P07996 homo sapien
25	224.5	11.0	1170	1 TSPI_MOUSE	P35441 mus musculus
26	223.5	10.9	824	1 AD08_HUMAN	P78325 homo sapien
27	222.5	10.9	1170	1 TSPI_BOVIN	Q28178 bos taurus
28	220	10.8	571	1 DISJ_BOVJA	P30431 bothrops ja
29	219.5	10.7	774	1 AD28_MOUSE	Q9jln6 mus musculus
30	214	10.5	813	1 AD33_HUMAN	Q9bz11 homo sapien
31	210.5	10.3	920	1 AD19_MOUSE	O35674 mus musculus
32	210	10.3	416	1 HR1B_TRIFL	P20164 trimeresuru
33	205.5	10.1	936	1 AD19_HUMAN	Q9h013 homo sapien

RESULT 1

ID	ATSL_MOUSE	STANDARD;	PRT;	968 AA.
DT	P97857; O54768;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).			
GN	ADAMTS1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RP	STRAIN=129/SVJ;			
RX	MEDLINE=98110583; PubMed=9441751;			
RA	Kuno K., Lizasa H., Ohno S., Matsushima K.;			
RT	"The exon/intron organization and chromosomal mapping of the mouse			
RT	ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";			
RL	Genomics 46:466-471(1997).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97150761; PubMed=8995297;			
RA	Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.,			
RA	Matsushima K.;			
RT	"Molecular cloning of a gene encoding a new type of metalloproteinase-			
RT	disintegrin family protein with thrombospondin motifs as an			
RT	inflammation associated gene.";			
RL	J. Biol. Chem. 272:556-562(1997).			
RN	[3]			
RP	CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.			
RX	MEDLINE=99303657; PubMed=10373500;			
RA	Kuno K., Terashima Y., Matsushima K.;			
RT	"ADAMTS-1 is an active metalloproteinase associated with the			
RT	extracellular matrix.";			
RL	J. Biol. Chem. 274:18821-18826(1999).			
RN	[4]			
RP	FUNCTION.			
RX	MEDLINE=20389568; PubMed=10930576;			
RA	Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.,			
RA	Ohno H., Matsushima K.;			
RT	"ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";			
RL	FEBS Lett. 478:241-245(2000).			
RN	[5]			
RP	FUNCTION, AND INDUCTION			
RX	MEDLINE=2043757; PubMed=10781075;			
RA	Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W.,			
RA	Richards J.S.;			
RT	"progesterone-regulated genes in the ovulation process: ADAMTS-1 and			
RT	cathepsin L proteases.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).			
CC	-!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE			
CC	INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY			
CC	SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH			

ALIGNMENTS

34	199	9.7	826	1	AD08_MOUSE	Q05910 mus musculus
35	197	9.6	413	1	ACIA_AKGAC	Q9pw35 agkistrodon
36	197	9.6	480	1	DISA_TRIGA	P15503 trimeresuru
37	197	9.6	797	1	AD33_MOUSE	Q923w9 mus musculus
38	196.5	9.6	814	1	AD15_HUMAN	Q13444 homo sapien
39	194	9.5	1074	1	SMSA_HUMAN	Q13591 homo sapien
40	192.5	9.4	1093	1	SMSB_MOUSE	Q60519 mus musculus
41	187.5	9.2	867	1	SSPO_BOVIN	P98167 bos taurus
42	186.5	9.1	1077	1	SMSA_MOUSE	Q62217 mus musculus
43	184	9.0	1173	1	TSPI_XENLA	P35448 xenopus lae
44	183.5	9.0	203	1	FIBR_AGKCO	P28891 agkistrodon
45	181.5	8.9	414	1	HRTD_CROAT	P15167 crotalus at

CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1691-GLU-1-LEU-1692
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX.
CC -1- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY
CC INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY
CC LIPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA
CC CELLS OF PREOVULATORY FOLLICLES.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 7.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB001735; BAA24501.1; ALT INIT.
DR EMBL; D67076; BAA11088.1; ALT_FRAME.
DR MEROPS; M12.222; -.
DR MGD; MGI:109249; Adamts1.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_prop.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTPeptidase.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_prop; 1.
DR SMART; SM00209; TSP1; 3.
DR PROSITE; PS02115; ADAM_MPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 48
FT PROPEP 49 253
FT CHAIN 254 968
FT SITE 206 206
FT METAL 402 402
FT METAL 403 403
FT ACT_SITE 406 406
FT METAL 412 412
FT METAL 477 559
FT DOMAIN 560 617
FT DOMAIN 618 725
FT DOMAIN 726 850
FT DOMAIN 851 909
FT DOMAIN 910 968
FT DOMAIN 195 199
FT DOMAIN 548 548
FT CARBOHYD 721 721
FT CARBOHYD 765 765
FT CARBOHYD 783 783
FT CARBOHYD 946 946
FT MUTAGEN 403 403
FT CONFLICT 335 335
FT CONFLICT 425 425
FT SEQUENCE 968 AA; 105841 MW; 42EBDA55499FB6C1 CRC64;
SQ

Query Match 65.5%; Score 1337; DB 1; Length 968;
Best Local Similarity 63.2%; Pred. No. 2e-103;
Matches 237; Conservative 51; Mismatches 81; Indels 6; Gaps 4;
QY 1 FVSPRYVETLVVADESMVKFAGDLEHYLLTLLATAARLYRHPISILNPINIVVKVLL 60
DB 254 FVSPRYVETLVVADESMVKFAGDLEHYLLTLLATAARLYRHPISILNPINIVVKVLLVI 313
QY 61 RDRSDGPKVTGNAALTRNFCANQKLNKVKYDKHPEYWDTAIFLTRODLGGATTCDTLGM 120
DB 314 YEEQKGPEVTSNAALTRNFCANQKLNKVKYDKHPEYWDTAIFLTRODLGGATTCDTLGM 373
QY 121 ADVCTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRANHMSP 180
DB 374 ADVCTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRANHMSP 433
QY 181 TLIQIDRANPWSACSAIITDLDGSHGDCILDPQSKPISLPEDLPQASTVLSQOCELAF 240
DB 434 MLSLIDHSQWSPSCSAIITDLDGSHGDCILDPQSKPISLPEDLPQASTVLSQOCELAF 493
QY 241 GVGSKPCP-YMQYCTKLWCTGKAKQWVCQTRHPFPAWDTSCGEGKCLKAGACVERHNLA 299
DB 494 GEESKHCPDAASTCTTLCCTGSGLLVCQTKHPFPAWDTSCGEGKCLKAGACVERHNLA 552
QY 300 KH--RVDSWAKWDPYPCSRCTCGGVQVLARRCTNPTPANGKYCEGVRYRSCNLE 356
DB 553 KHFPATPVHSGWPGWPGWDCSRTCGGVQVYTMRECDNPVKNKGKYCEGVRYRSCNIE 612
QY 357 PCPSASGKSPREQ 371
DB 613 DCPDN-NGKTPREQ 626

RESULT 2
ATSL_HUMAN STANDARD; 967 AA.
ID Q9UHI8; Q9UP80; Q9UH83; Q9P2K0; Q9NSJ8;
AC 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).
GN ADAMTS1 OR METH1 OR KIAA1346.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Casas C., Pritchard M.A., Estivill X., Arbones M.L.;
RT "Cloning, characterization and mapping on human chromosome 21 of the
RT orthologue of murine Adamts-1";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC TISSUE=Heart;
RX MEDLINE=99367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
RA Lombardo M., Iruela-Arispe M.L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
RT family of proteins with angio-inhibitory activity";
RL J. Biol. Chem. 274:23349-23357(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Endothelial cells;
RX MEDLINE=20247184; PubMed=10785405;
RA Glienke J., Schmitt A.O., Filarsky C., Hinzmann B., Weiss B.,
RA Rosenthal A., Thierach K.H.;
RT "Differential gene expression by endothelial cells in distinct
RT angiogenic states";
RL Eur. J. Biochem. 267:2820-2830(2000).
RN [4]
RP SEQUENCE FROM N.A.

RC TISSUE=Brain;
 RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes.XVI.
 The complete sequences of 150 new cDNA clones from brain which code
 for large proteins in vitro.";
 RL DNA Res. 7:65-73(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
 RA Ohki M., Takagi T., Sakai K., Taudien S., Blechschmidt K., Polley A.,
 RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordtsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Riesselmann L., Dagand E.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.,
 RA "The DNA sequence of human chromosome 21.";
 RT Nature 405:311-319(2000).
 RN [6]
 RP SEQUENCE OF 418-967 FROM N.A.
 RC TISSUE=Melanoma;
 RX MEDLINE=20289799; PubMed=10830953;
 RA Blum H., Bauersachs S., Mewes H.-W., Weil B., Wiemann S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
 INVOLVED IN ITS TURNOVER (BY SIMILARITY). HAS ANGIOGENIC INHIBITOR
 ACTIVITY. ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH
 VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
 CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE.
 CC -!- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1938-GLU-|-LEU-1939
 SITE. WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 matrix (by similarity).
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
 SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF170084; AAF15317.1; -;
 DR EMBL: AF060152; AAD48080.1; ALT_INIT.
 DR EMBL: AF207664; AAF23772.1; -;
 DR EMBL: AB037767; BAA92584.1; ALT_INIT.
 DR EMBL: AF001697; BAA95502.1; -;
 DR EMBL: AL162080; CAB82413.1; -;
 DR MEROPS: M12.222; -;
 DR Genew: HGNC:217; ADAMTS1.
 DR MIM: 605174; -;
 DR InterPro: IPR001762; Disintegrin.
 DR InterPro: IPR002870; Pep_M12B_propep.
 DR InterPro: IPR001590; Reprolysin.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR000130; Zn_MTPeptase.
 DR Pfam: PF00090; tsp.1; 6.
 DR Pfam: PF01421; Reprolysin; 2.
 DR Pfam: PF01562; Pep_M12B_propep; 2.
 DR SMART: SM00209; TSP1; 3.

DR PROSITE; PS50215; ADAM_MEPRO; 1.
 DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 DR PROSITE; PS50092; TSP1; 2.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Extracellular matrix; Heparin-binding.
 FT SIGNAL 1 49
 FT PROPEP 50 252
 FT CHAIN 253 967
 FT SITE 198 198
 FT METAL 401 401
 FT ACT_SITE 402 402
 FT METAL 405 405
 FT METAL 411 411
 FT DOMAIN 476 559
 FT DOMAIN 560 616
 FT DOMAIN 617 724
 FT DOMAIN 725 849
 FT DOMAIN 850 908
 FT DOMAIN 909 967
 FT DOMAIN 843 846
 FT CARBOHYD 547 547
 FT CARBOHYD 720 720
 FT CARBOHYD 764 764
 FT CONFLICT 227 227
 FT CONFLICT 468 468
 FT CONFLICT 561 561
 FT SEQUENCE 967 AA; 105383 MW; C189389324741ED1 CRC64;
 SQ
 Query Match 65.3%; Score 1334; DB 1; Length 967;
 Best Local Similarity 63.5%; Pred No. 3.5e-103;
 Matches 238; Conservative 50; Mismatches 81; Indels 6; Gaps 4;
 QY 1 FVSIPIRYVETLVVADBSMYKFGADLEHYLLTLATAARLYRHPSLNPNIVVVKVLL 60
 DB 253 FVSSHYRYVETLVVADBSMYKFGADLEHYLLTLATAARLYRHPSLNPNIVVVKV 312
 QY 61 RDRDSGPKVTGNAALTRNFCWOKKLVSKDHPYWDYTAFLTRQDLGGATCDDTGM 120
 DB 313 HDEQKGEVTSNAALTRNFCWOKKLVSKDHPYWDYTAFLTRQDLGGATCDDTGM 372
 QY 121 ADVGTCMDPKRSCVIEDDGLPSAFTTAHELGHVFNHMDNVKVEEYFGKLRANHMSP 180
 DB 373 ADVGTCMDPKRSCVIEDDGLPSAFTTAHELGHVFNHMDNVKVEEYFGKLRANHMSP 432
 QY 181 TLIQIDRANPWACSAIITDLDSDHGDCLDQSPISLPEDLPGASYTLSSQCELA 240
 DB 433 MLSNLDHSDQWSPSCSAIITDLDSDHGDCLDQSPISLPEDLPGASYTLSSQCELA 492
 QY 241 GVGSKPCP-YMOYCTKLTCTGKAGOMVCTRHFPWADTSGEGKGLKLGACVHRNLN 299
 DB 493 GEDSKHCPDAASTCTLWCTGTSGGVLCQTKHFPWADTSGEGKGLKLGACVHRNLN 551
 QY 300 KH---RVDSWAKWDPYPCSRCTCGGVQLARRQCTNPANGKGYCEGVRYRSCNLE 356
 DB 552 KHFDTPFHGSGWGMGPGWGDSCRTCGGVQYTMRECDNPVKNKGYCEGVRYRSCNLE 611
 QY 357 PCPSSASGKSPREEQ 371
 DB 612 DCPDN-NGKTFREEQ 625
 RESULT 3
 ATSL_RAT
 ID ATSL_RAT STANDARD; PRT; 967 AA.
 AC Q9WUQ1; Q9ER11;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
 GN ADAMTS1.
 OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RA Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,
RA Little S.P.;
RT "Induction of a disintegrin and metalloprotease with the
thrombospondin type I motif (ADAMTS).";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 18-967 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=20304099; PubMed=10847486;
RA Diamantis I., Luechi M., Hoesli M., Reichen J.;
RT "Cloning of the rat ADAMTS-1 gene and its down regulation in
endothelial cells in cirrhotic rats.";
RL Liver 20:165-172(2000).
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU-1-LEU-1684
SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (by similarity).
CC -1- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM
CIRRHOTIC LIVER.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
DR EMBL; AF149118; AAD34012.1; -;
DR EMBL; AF304446; AAG29823.1; -;
DR MEROPS; M12.222; -;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_M12Bptdase.
DR Pfam; PF00090; tsp_1; 6.
DR Pfam; PF01421; Reprolysin; 2.
DR Pfam; PF01562; Pep_M12B_propep; 2.
DR SMART; SM00209; TSP1; 3.
DR PROSITE; PS02015; ADAM_MEP13; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS50092; TSP1; 2.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 54
FT PROPEP 55 252
FT CHAIN 253 967
FT SITE 205 205
FT METAL 401 401
FT ACT_SITE 402 402
FT METAL 405 405
FT ZINC (CATALYTIC) (BY SIMILARITY).
FT Cysteine SWITCH (POTENTIAL).
FT ZINC (CATALYTIC) (BY SIMILARITY).
FT ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 411
FT DOMAIN 476 558
FT TSP TYPE-1 1.
FT DOMAIN 559 615
FT DOMAIN 616 724
FT DOMAIN 725 857
FT DOMAIN 858 907
FT DOMAIN 908 967
FT DOMAIN 908 967
FT DOMAIN 194 198
FT CARBOHYD 547 547
FT CARBOHYD 720 720
FT CARBOHYD 764 764
FT CARBOHYD 782 782
FT CARBOHYD 945 945
FT CONFLICT 21 21
FT CONFLICT 26 31
FT CONFLICT 49 49
FT CONFLICT 72 72
FT CONFLICT 79 79
FT CONFLICT 249 249
FT CONFLICT 262 265
FT CONFLICT 607 607
FT CONFLICT 936 936
FT CONFLICT 962 962
SQ SEQUENCE 967 AA; 105705 MW; F93G864F6DCDB4CF CRC64;
Query Match 65.0%; Score 1327; DB 1; Length 967;
Best Local Similarity 62.7%; Pred. No. 1.3e-102;
Matches 235; Conservative 52; Mismatches 82; Indels 6; Gaps 4;
QY 1 FVSIPRYVETLVVADSMVKFHGADLEHYLLTLATAARLYRHPSILNPINIVVVKVLL 60
DB 253 FVSSPRYVETLMVADQSMADFHGSLKHYLLTLFSAARFYKHPHSIRNSISLVVVKI 312
QY 61 RDRSGPKVTGNALTLNFCQWKKLNKVSXKHPYWDTAILFTRODLGCATCTDGLM 120
DB 313 YEEQKGEVTSNALTLNFCQWKKLNKVSXKHPYWDTAILFTRODLGCATCTDGLM 372
QY 121 ADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVNMFDNKKVCEEVFGKLRANHMSP 180
DB 373 ADVGTCDPSRSCSVIEDDGLQAAFTTAHELGHVNMFDNKKVCEEVFGKLRANHMSP 432
QY 181 TLIIQDRANPWSACSAIITDLDGSHGCDLDDQSPKISLPEDLPQASYSYLSQCELA 240
DB 433 MLSSLDHSPWSPCSAYMVVTFSLDNGHGECLMDKPNPKLPDLPGLTLDANRQCQTF 492
QY 241 GVGSKPCP-YMYCYCKLWCTGKAKQWYQTRHFWADGTSCGEGKLCIGKACVHRHNL 299
DB 493 GEESTHCDDAATCTSTLWCTGTSGLLVQCTKHFPWADGTSCGEGKWCYSGKCVNKTDM- 551
QY 300 KH--RVDGSAKWDPPYPCSRCTCGGVQALRRQCTNPTPANGKYCEGVRYKYSCHLE 356
DB 552 KHFAFPVHSGWPGWPGWPGWPGWPGWPGWPGWPGWPGWPGWPGWPGWPGWPGWPG 611
QY 357 PCPSSASGSKSFREEQ 371
DB 612 DCPDN-NGKTFREEQ 625
RESULT 4
ATS4_HUMAN STANDARD; PRT; 837 AA.
ID AC 075173; Q9UN83;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
DE (ADMP-1).
GN ADAMTS4 OR KIAA0688.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

RN RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98403880; PubMed=9734811;
 RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
 RA Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. X.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:169-176(1998).
 RN [2]
 RN RP SEQUENCE FROM N.A.
 RX MEDLINE=9286303; PubMed=10356395;
 RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
 RA Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
 RA Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,
 RA Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
 RA Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
 RA Trzaskos J.M., Arner E.C.;
 RT "Purification and cloning of aggrecanase-1: a member of the ADAMTS
 RT family of proteases.";
 RL Science 284:1664-1666(1999).
 RN [3]
 RN RP SEQUENCE FROM N.A.
 RA Sawaji Y., Nagase H., Saklatvala J., Clark A.R.;
 RT "ADAMTS-4 genomic locus.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RX MEDLINE=20400518; PubMed=10827174;
 RA Tortorella M.D., Pratta M.A., Liu R.-Q., Abbaszade I., Ross H.,
 RA Burn T.C., Arner E.C.;
 RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
 RT aggrecan substrate recognition and cleavage.";
 RL J. Biol. Chem. 275:25791-25797(2000).
 CC -1- INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
 CC DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE
 CC CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN
 CC ALZHEIMER'S DISEASE.
 CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393
 CC site.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: secreted. Associated with the extracellular
 CC matrix (by similarity).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED
 CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
 CC -1- INDUCTION: BY INTERLEUKIN-1.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
 CC -1- CAUTION: HAS SOMETIMES BEEN REFERRED TO AS ADAMTS2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB014588; BAA31631.1; -;
 CC EMBL; AF148213; AAD41494.1; -;
 CC EMBL; AY044847; AAL02262.1; -;
 CC MEROPS; M12.221; -;
 CC Genew; HGNC:220; ADAMTS4.
 CC MIM; 603876; -;
 CC InterPro; IPR001762; Disintegrin.
 CC InterPro; IPR001590; Reprolysin.
 CC InterPro; IPR000884; TSP1.

DR InterPro; IPR000130; zn_MTpeptdse.
 DR Pfam; PF00090; tsp_1; 1.
 DR Pfam; PF01421; Reprolysin; 1.
 DR SMART; SM00209; TSP1; 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR PROSITE; PS00215; ADAM_MEPRO; 1.
 DR PROSITE; PS00092; TSP1; 1.
 DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Extracellular matrix.
 FT SIGNAL 1 51 POTENTIAL.
 FT PROPEP 52 212
 FT CHAIN 213 837 ADAMTS-4.
 FT SITE 194 194 CYSTEINE SWITCH (POTENTIAL).
 FT METAL 361 361 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 362 362 BY SIMILARITY.
 FT METAL 365 365 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 371 371 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 437 519 DISINTEGRIN-LIKE.
 FT DOMAIN 520 576 TSP TYPE-1.
 FT DOMAIN 577 685 SPACER.
 FT DOMAIN 686 837 CYS-RICH.
 FT DOMAIN 247 252 POLY-ALA.
 FT CARBOHYD 68 68 N-LINKED (GLCNAC... (POTENTIAL).
 FT CONFLICT 77 77 A -> T (IN REF. 1).
 FT CONFLICT 626 626 R -> Q (IN REF. 3).
 FT CONFLICT 682 682 G -> R (IN REF. 3).
 SQ SEQUENCE 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;
 Query Match 57.9%; Score 1182.5; DB 1; Length 837;
 Best Local Similarity 55.5%; Pred. No. 1.2e-90;
 Matches 208; Conservative 56; Mismatches 106; Indels 5; Gaps 4;
 Qy 1 FVSIPRIVETLVVADESMVFHGDLEHYLLTLTAARLYRHPISILNINIVVKVLL 60
 Db 213 FASLSRFVETLVVADDDKMAAFHAGLAKRYLLTVAATAAKAFKHPISIRNPVSLVTVTLVL 272
 Qy 61 RDRDSGPKVTGNAALTIRNECAMQKLNKYSKHPYWDTAIFLTRQDLGATTCDTLGM 120
 Db 273 GSGEQPGQVPSAAQTLRSFCAMQGLNTPEDSDPDHFDTAIFLTRQDLGATTCDTLGM 332
 Qy 121 ADVGTMCDPKSCSVIEDGLPSAFTTAHELGHVFNPHDNVKNVCEVEVKL-RANHMS 179
 Db 333 ADVGTVCDPARSALVEDDGLQSAFTAHELGHVFNHMDNSRPSISLNGPLSTSRHWA 392
 Qy 180 PTLIQIDRANPWSACSAIITDFLDGSGHGLDLPQSKPISLPELPGASYTILSOCELA 239
 Db 393 PVMAHVDPPEPSPCSARFITTDFLDNGYHCLLDKPEAPLHPVTPFGKYDADRQCQLT 452
 Qy 240 FGVSCKPCPYM-QYCTKLWCTGKAKGMVCTRHFPWADTSCGEGKLCGLKACVERHNL 298
 Db 453 FGPOSRRHCPQLPPPCALWCSGHLNGHAMCOTRHSPWADTGPCGPAQACGGHCLHMDQL 512
 Qy 299 NKHRVD--GSWAKWDPYGPCSRTCGGQVQLARRQCTNPTPANGKCYCEGVYRYKRSNLE 356
 Db 513 QDFNIPQAGGWPWGWCDCSRCTCGGQVQFSRRDCTRPVPRNGKCYCEGRRTFRSCNTE 572
 Qy 357 PCPSASCKSPREEQ 371
 Db 573 DCP-TGSALTFREEQ 586
 RESULT 5
 ATSA_RAT STANDARD; PRT; 630 AA.
 AC Q9ESP7; Q9ESP8; Q9ESP6;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
 DE (Fragment).
 GN ADAMTS4.

QY	1	FVSIPIRYVETLVVADSMVKFKGADLEHYLLFLATAARLYRHPSSLNPINLVVVKVLL	80
Db	6	PASLSRFVETLVVADDDKMAAFHGAGLKHLLVTMAAAAKAFKHPSTRPNVNVVTRLVIL	65
QY	61	RDRDSGPKVTGNAALTLRNFCAWOKLKNVSKDHPYWDFTAILFTRODLGCAATTCDTLGM	120
Db	66	GSQGEYVPGVQPSAAQTLRFSCTWOKGNPNPNDSDPDHFTAILFTRODLGCVSTCDALGM	125
QY	121	ADYGTMCDDPKRSVTDGGLSAFTTAHELGHVFNPHNDNVKVCBEVEFGK-LRANHIMS	179
Db	126	AGVGTVCDPARSCAIYEDDGLQSAFTAHELGHVFNMLHNSKPCANLNGQSSSRHVMA	185
QY	180	PTLIQIDRANPWSACSAAIITFDLSGHGDCLLDQPSKPISLPEDLPGASYTLSSQCELA	239
Db	186	PVMAHYDPEEPWSPCSARFTFDLDNGYGHCLLDKPEAPLHLPVTPPGKDYDADROCOLT	245
QY	240	FGVGSKPCPYM-QYCTKLJCTGKAKGOMVCOTRHPFWADGTCGEGKLCCLKGACVERHNL	298
Db	246	FGPDSHSCPOLPPPCALCMFCFHLNHHAMCQTKHSPWADGTPCGPAQACMGRCCLHVDQL	305
QY	299	NKHRVD--GSWAKWDPIPGCSRTCCGGVQLARRQCTNPTPANGGKYCGEVVKYKRCNLE	356
Db	306	KDENIPQAGCGWPGWPGDCSRTCCGGGVQFSRDCTKPVPRNGGKYCEGRRTPFRSCNTK	365
QY	357	PCPSSASGKSFREEQ 371	
Db	366	NCP-HGSALTFREEQ 379	
RESULT 6			
ID	ATS8_MOUSE	STANDARD;	PRT; 905 AA.
AC	P57110;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase		
DE	with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2).		
GN	ADAMTS8.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
[I]			
FN	SEQUENCE FROM N.A.		
RP	MEDLINE=200791168; PubMed=10610729;		
RX	Georgiadis K.E., Hirohata S., Seidlin M.F., Apte S.S.;		
RA	"ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on		
RT	mouse chromosome 9 and human chromosome 11.";		
RT	Genomics 62:312-315(1999).		
RL			
CC	-!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES (BY SIMILARITY).		
CC	-!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).		
CC	-!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular		
CC	matrix (by similarity).		
CC	-!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT LUNG AND HEART		
CC	AND LOW EXPRESSION DURING MOUSE DEVELOPMENT.		
CC	-!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT		
CC	FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX		
CC	-!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY		
CC	SIMILARITY).		
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.		
CC	-!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.		
CC	-!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AF175282; AAF25805.1; -		


```
DR HSP: P34179; 1IAG.
DR MEROPS; M12.226; -.
DR MGD; MGI:1353468; AdamtS8.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTPeptide.
DR Pfam; PF00090; tsp_1; 2.
DR Pfam; PF01421; Reprolysin; 1.
DR SMART; SM00209; TSPl; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS0092; TSPl; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 28
FT PROPEP 29 228
FT CHAIN 229 905
FT METAL 378 378
FT ACT_SITE 379 379
FT METAL 382 382
FT METAL 388 388
FT DOMAIN 453 541
FT DOMAIN 542 598
FT DOMAIN 599 705
FT DOMAIN 706 847
FT DOMAIN 848 905
FT CARBOHYD 415 415
FT CARBOHYD 480 480
FT CARBOHYD 506 506
FT CARBOHYD 615 615
SQ SEQUENCE 905 AA; 98879 MW; 124D4132B33A0CAE CRC64;

Query Match 55.2%; Score 1128; DB 1; Length 905;
Best Local Similarity 53.8%; Pred. No. 4.3e-86;
Matches 205; Conservative 64; Mismatches 100; Indels 12; Gaps 6;

QY 1 FVSIPRYVETLVVADESVMKFGADLEHVLTLATAARLVRHPSILNPINIVVKVLL 60
Db 229 FVSEARFVETLLVADASMAAFVGTDLQNHILTVMSAARIKHPISRVNLVVKVLI 288
QY 61 RDRSGPKVTGNAALTRNFCWAKKLNKVSDEYDWTALFTFTRDLCG-ATTCDTLG 119
Db 289 EKERMGPVSDNGGLTRNFCWSORRPNKPSDRHPHYDTALFTFRNFCGKGEQCDTLG 348
QY 120 MADVGTCDPKSCSVIEDGLPSAFTAHGLGHVFNPHDNVNVKCEVFGLRANHMS 179
Db 349 MADVGTCDPKSCSVIKDEGLQAAYTLAHLGHVLSMPHDDSKPCVRLFGPMGKYHMA 408
QY 180 PTLIQIDRANPWSAGSAALITDFLDGHDCLLDQPSKPSIPEDLPGLAS--YTLSSQCE 237
Db 409 PFFIHVNTLPWSPCSAVYLTLELDDGHDCLLDAPTSVLPVLTGPGHSTLYELDQCK 468
QY 238 LAFVGSRKPCP---YMQYCTKLWCTGAKGQMVQCOTRH--FPWADGTSCGEGKLCGLKAC 292
Db 469 QIFGPDFRHCPTNSVEDICVOLCAHRDSDPEICTKNGSLIWDAGTGPCGHLCLDGSC 528
QY 293 VERHNLN--KHRVDSWAKWDYPCSRCTCGGVQLARQCTNPTTPANGKYCEGVRYKY 350
Db 529 VLKEDVENPKAVDWDGFWRPWGQCSRTCGGIGQFSNRECDNPMQNGRFLGERVYKY 588
QY 351 RSCNLEPCPSSASGSKSFREQ 371
Db 589 QSCNTEECPP--FNGKSFREQ 607

RESULT 7
AT88 HUMAN
ID AT88 HUMAN STANDARD; PRT; 890 AA.
AC Q9UP79; Q9NZ50;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
```

```
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2)
DE (METH-8).
GN ADAMTS8 OR METH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=99367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
RA Lombardo M., Iruela-Arispe M.L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
RT family of proteins with angio-inhibitory activity.";
RL J. Biol. Chem. 274:23349-23357(1999).
RN [2]
RP SEQUENCE OF 195-440 FROM N.A.
RX MEDLINE=20079168; PubMed=10610729;
RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on
RT mouse chromosome 9 and human chromosome 11.";
RL Genomics 62:312-315(1999).
CC -1- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER
CC EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND
CC KIDNEY.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF060153; AAD48081.1; -.
CC EMBL; AF175283; AAP25806.1; -.
CC HSP: P34179; 1IAG.
CC MEROPS; M12.226; -.
CC Genew; HGNC:224; ADAMTS8.
CC MIM; 605175; -.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; Pep_M12B_propep.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSPl.
CC InterPro; IPR000130; Zn_MTPeptide.
CC Pfam; PF00090; tsp_1; 2.
CC Pfam; PF01421; Reprolysin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC SMART; SM00209; TSPl; 2.
CC PROSITE; PS0215; ADAM_MEPRO; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS0092; TSPl; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 27
FT PROPEP 28 214
FT CHAIN 215 890
FT METAL 364 364
FT ZINC (CATALYTIC) (BY SIMILARITY).
```

```
FT ACT_SITE 365 365 BY SIMILARITY.
FT METAL 368 368 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 374 374 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 439 526 TSP TYPE-1 1.
FT DOMAIN 527 583 TSP TYPE-1 1.
FT DOMAIN 584 690 CYS-RICH.
FT DOMAIN 691 832 SPACER.
FT DOMAIN 833 890 TSP TYPE-1 2.
FT DOMAIN 890 900 POLY-PRO.
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 600 600 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 195 195 E -> R (IN REF. 2).
FT CONFLICT 413 440 YTELLDGGHGVCLDADGALPLPTGL -> FSGCHLQGW
FT SEQUENCE 890 AA; 96671 MW; 57D70E03D5739D3 CRC64;

Query Match 55.2%; Score 1126.5; DB 1; Length 890;
Best Local Similarity 53.3%; Pred. No. 5.6e-86;
Matches 203; Conservative 64; Mismatches 101; Indels 13; Gaps 7;

QY 1 FVSIPRYVETLVVADESMVKFHGADLEHYLLTLLATAARLYRHPSTLNEINIVVVKVLL 60
DB 215 FVSEARFVETLVVADASMAAFYGADJONHILTLMSVAARIYKHPSTKSNLNMVVKVLLV 274
QY 61 RDRDSGPKVTGNALTRNFCAMQKLNKYNKHPYDWTALFTQDLCGAT-TCDTLGL 119
DB 275 EDEKNGPEVSDNGLTRNFCAMQKLNKYNKHPYDWTALFTQDLCGAT-TCDTLGL 334
QY 120 MADVGMCDPKRSCVIEDDGLPSATTAHELGHVFNPHNDVNVKVEEFGKLRNHHMS 179
DB 335 VADITGTCIDPNKSCVIEDDGLPSATTAHELGHVFNPHNDVNVKVEEFGKLRNHHMS 394
QY 180 PTLQIDRANPWSACSAITTDPLDSDGDCDLDQPSKPLSLPDLPG--ASYTLSQOCE 237
DB 395 PLFVHLNQTLPWSPCSAMYTELDDGGHGDCLDADGALPLTGLPGRMALYQLDQOCR 454
QY 238 LAFVGSGKPCP---YMOYCTKLMCTGKAKGMVQCOTRH--FPWADGTCGEGKCLKLGAC 292
DB 455 QIFGPDPRHCPNTSAQDVCAQLWCHTDG-AEPLCHTKNGSLPWADGTPCGPGHLCSEGC 513
QY 293 VERHNLNKR--VDGSHAKWDPGCSRTCCGGVQVQLARQCTNPTANGKICEGVRYKY 350
DB 514 LPEEVEERPKPVVDGGWAPGPGWEGECSTCGGGVQFVSHRECKDPEPONGRGYCLGRRAKY 573
QY 351 RSCNLEPCPSSAGSKSPREQ 371
DB 574 QSCHEBEP--PDGKSPREQ 592

RESULT 8
ID ATSS_MOUSE STANDARD; PRT; 930 AA.
AC Q8R001;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
DE (ADMP-2) (Implantin).
GN ADAMTS5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99395124; PubMed=10464288;
RA Hurskainen T.L., Hirohata S., Seidlin M.F., Apte S.S.;
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RT zinc metalloproteinases.";
```

```
J. Biol. Chem. 274:25555-25563(1999).
-|- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
-|- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393 site.
-|- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
-|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (by similarity).
-|- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PERI-IMPLANTATION PERIOD IN EMBRYO AND TROPHOBLAST AND AT LOW OR UNDETECTABLE LEVEL THEREAFTER.
-|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
-|- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY SIMILARITY).
-|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
-|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
-|- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; AF140673; AAD56356.1; -.
MEROPS; M12.225; -.
MGD; MGI:1346321; Adamts5.
InterPro; IPR001762; Disintegrin.
InterPro; IPR002870; Pep_M12B_propep.
InterPro; IPR001590; Reprolysin.
InterPro; IPR000884; TSP1.
InterPro; IPR000130; zn_Mtpeptidse.
Pfam; PF00090; Tsp1; 2.
Pfam; PF01421; Reprolysin; 1.
Pfam; PF01562; Pep_M12B_propep; 1.
SMART; SM00209; TSP1; 2.
PROSITE; PS00215; ADAM_MEPRO; 1.
PROSITE; PS00142; ZINC_PROTEASE; 1.
PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 261 POTENTIAL.
FT CHAIN 262 930 ADAMTS-5.
FT SITE 209 209 CYSTEINE SWITCH (POTENTIAL).
FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 411 411 BY SIMILARITY.
FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 485 566 DISINTEGRIN-LIKE.
FT DOMAIN 567 623 TSP TYPE-1 1.
FT DOMAIN 624 731 CYS-RICH.
FT DOMAIN 732 874 SPACER.
FT DOMAIN 875 930 TSP TYPE-1 2.
FT DOMAIN 930 930 POLY-ALA.
FT DOMAIN 930 930 POLY-ARG.
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 930 AA; 101780 MW; 84DE84B26170D4DC CRC64;

Query Match 54.6%; Score 1114; DB 1; Length 930;
Best Local Similarity 53.6%; Pred. No. 6.4e-85;
Matches 201; Conservative 58; Mismatches 106; Indels 10; Gaps 3;

QY 2 VSPRYVETLVVADESMVKFHGADLEHYLLTLLATAARLYRHPSTLNEINIVVVKVLLR 61
```

Db 263 ISRAQVELLVADSSMARMYGRGLQHYLLTLASIANFLYSHASHENHRLAVKVVVLT 322
 QY 62 DRDSPKVTGNAULTLRNFCNAQKLNKVKSDKHPEYWDYTAIFTRQDLCCGATTCDTLGM 121
 Db 323 DKDTSLVYSKNAATTLKFNCKWQHNGHNGDHEHYDAAALFTREDLGHSCDTLGM 382
 QY 122 DVGTMCDPKRSCSVIEDDGLPSAFTTAHELGLHFNPHDNVKNVCEEVFGKRLRHHMSPT 181
 Db 393 DVGITCSPERSCAVIEDDGLHAAFTVAHEIGHLLGLSHDSKFCFEEFNFTEDKRLMSI 442
 QY 182 LIQDRANPWSACSAIITDFLDSCGHGDCLLDQSPKISLPEDLPGASYTLISQOCELAF 241
 Db 443 LTSIDASKPWSKTSATITEFLDDHGNCLLDLPRKQLGLPEELPGOTYDQOCNLTFG 502
 QY 242 VGSRCPCYMYCTKLWCTGKAGQWQVQTRFPWADGTSGEGKLCUKGACVER-----H 296
 Db 503 PEYSVCPGDMCARLWCNAVVRQGMVCLTKKLPAVEGTCGKRVCLQKGVCKDKTKKYY 562
 QY 297 NLNKHVRDGSWAKWDPYGPCSRTCGGVQVLARRQCTNPTPANGGKYCEGVKVRKNSCLE 356
 Db 563 STSSH---GNWGSNGPWGQCSRGSGGVQVFPAYRHCNPNPNSRGYCTGKRAIYRSCVT 619
 QY 357 PCPSASGSPREQ 371
 Db 620 PCP--PNGKSPREQ 632

RESULT 9

ID ATSS_HUMAN STANDARD; PRT; 930 AA.
 AC Q9UNAO; Q9URP2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
 DE with chromospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2)
 DE (ADMP-2) (ADAM-TS 11).
 GN ADAMTS5 OR ADMP2 OR ADAMTS11.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=9367476; PubMed=10438522;
 RA Abbaszade I., Liu R.-Q., Yang F., Rosenfeld S.A., Ross O.H.,
 RA Link J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M.,
 RA Wynn R., Duke J.L., George H.J., Hillman M.C. Jr., Murphy K.,
 RA Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner R., Nagase H.,
 RA Ito Y., Newton R.C., Magolda R.L., Trzaskos J.M., Hollis G.F.,
 RA Arner E.C., Burn T.C.;
 RA "Cloning and characterization of ADAMTS11, an aggrecanase from the
 RA ADAMTS family.";
 RL J. Biol. Chem. 274:23443-23450(1999).
 [2]
 RN SEQUENCE FROM N.A.
 RC MEDLINE=2089799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
 RA Ohki M., Takagi T., Sakaki Y., Kumf K., Lehmann R., Patterson D.,
 RA Menzel U., Delabar J., Schillhabel M., Schudy A., Zimmermann W.,
 RA Reichwald K., Rump A., Shibuya K., Kawasaki K., Asakawa S.,
 RA Rosenthal A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).

RN [3]
 RP SEQUENCE OF 413-930 FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=99395124; PubMed=10464288;
 RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
 RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
 RT zinc metalloproteases.";
 RL J. Biol. Chem. 274:25555-25563(1999).
 CC -!- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
 CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
 CC DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
 CC PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
 CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393
 CC site.
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA
 CC BUT ALSO IN OTHER TISSUES, SUCH AS HEART AND BRAIN, AND ALSO
 CC CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE,
 CC CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN
 CC ARTHRITIC PATIENT.
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
 CC SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF142099; AAD49577.1; -;
 DR EMBL: AF001698; BAA95504.1; -;
 DR EMBL: AF001697; BAA95503.1; -;
 DR EMBL: AF141293; AAF02493.1; -;
 DR HSP: Q9PW35; 1BUD.
 DR MEROPS: M12.225; -;
 DR Genew: HGNC:221; ADAMTS5.
 DR MIM: 605007; -;
 DR InterPro: IPR001762; Disintegrin.
 DR InterPro: IPR002870; Pep_M12B_propep.
 DR InterPro: IPR001590; Reprolysin.
 DR InterPro: IPR000884; TSP1.
 DR InterPro: IPR000130; Zn_MTpeptdse.
 DR Pfam: PF00090; tsp_1; 2.
 DR Pfam: PF01421; Reprolysin; 1.
 DR Pfam: PF01562; Pep_M12B_propep; 1.
 DR SMART: SM00209; TSP1; 2.
 DR PROSITE: PS50215; ADAM_MPRO; 1.
 DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
 DR PROSITE: PS50092; TSP1; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Extracellular matrix.
 FT SIGNAL 1 16
 FT PROPEP 17 261 POTENTIAL.
 FT CHAIN 262 930 ADAMTS-5.
 FT SITE 209 209 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 410 411 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 411 411 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 485 566 DISINTEGRIN-LIKE.
 FT DOMAIN 567 623 TSP TYPE-1 1.
 FT DOMAIN 624 731 CYS-RICH.
 FT DOMAIN 732 874 SPACER.

```
FT DOMAIN 875 930 TSP TYPE-1 2.
FT DOMAIN 37 41 POLY-ALA.
FT DOMAIN 257 261 POLY-ARG.
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 138 138 A -> G (IN REF. 2).
FT CONFLICT 614 614 R -> H (IN REF. 3).
FT CONFLICT 692 692 P -> L (IN REF. 2).
SQ SEQUENCE 930 AA; 101715 MW; B64281502F28193B CRC64;

Query Match 54.4%; Score 1110; DB 1; Length 930;
Best Local Similarity 53.6%; Pred. No. 1.4e-84;
Matches 201; Conservative 56; Mismatches 108; Indels 10; Gaps 3;

QY 2 VSPRYVETLVVADESVMKPHGADLEHYLLTLTAARLYRHPSTIINPINIVVVKVLLLR 61
DQ 263 ISRARQVELLVADASMARLYGRGLQHYLLTLASIANRLYSHASIEHRLAVKVVVLG 322
QY 62 DRSGPKVTGNAALTNRFCWQKLNKYSKHPYWDYTAIFTRDLCGATTCDDTLGMA 121
DQ 323 DKKSLEVSNAATLTKFKCKWQHQNQLGDDHEHYDAALFTREDLCGHHSDFLGM 382
QY 122 DVGTMDPKRSCSVIEDGLPSAFTTAHELGHVFNPHDNKVKCEVFGKLRANHHMSPT 181
DQ 383 DVGITCSPERSCAVIEDDGLHAAFTVAHEIGHELGLSHDSDKCEETFGSTEDKRLMSSI 442
QY 182 LIQIDRANPHSACSAIITDFLDSGHDCDCLLDQSPKISIPEDLPACASTLSOOCELANG 241
DQ 443 LTSIDASKPKSKTSAPITEFDDGHGNCCLLDLPKQILGPEELPGQTYDATQCCNLTFG 502
QY 242 VGSKPCPYMYCKTLGAKQGMVQCTRHFPWADTSCGEGKCLKLGACVER-----H 296
DQ 503 PEYSVCFGMVDCARLWCAVYRQGMVCLTKKLPVAVSTPCGKGRICLQKGVCKTKKYY 562
QY 297 NLNKHVRDGSWAKWDPYPCSPRCGCGVQLARRQCTNPTPANGKKGCEGVYKRYSCNLE 356
DQ 563 STSSH---GNWGSNGWGQSCRS CGGVQFAYRHCCNNPAPRNNRGYCTGKRAIYRSCSLM 619
QY 357 PCPSSASGKGFREQ 371
DQ 620 PCP--PNGKSRHEQ 632

RESULT 10
ID AT9_HUMAN STANDARD; PRT; 1629 AA.
AC Q9P2N4; Q9NR29;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2002 (Rel. 40, Last sequence update)
DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).
GN ADAMTS9 OR KIAA1312.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE=Petal;
RX MEDLINE=203961138; PubMed=10936055;
RA Clark M.E., Kellner G.S., Turbeville L.A., Boyer A., Arden K.A.,
RA Maki R.A.;
RT "ADAMTS 9, a novel member of the ADAM-TS/Metallospodin gene
RT family.";
RL Genomics 67:343-350(2000).
RN [2]
RP SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
```

```
RT RT
RT RT
RT RT
RL DNA Res. 7:65-73(2000).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES. EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG, PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR THYMUS.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration with the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AF261918; AAF89106.1; -.
DR EMBL; AB037733; BAA92550.1; -.
DR HSP; P15167; IATL.
DR MEROPS; M12.021; -.
DR Genew; HGNC:13202; ADAMTS9.
DR MIM; 605421; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_Mtpeptidse.
DR Pfam; PF00090; tsp_1; 11.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 12.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS00092; TSP1; 9.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen; Repeat; Extracellular matrix; Alternative splicing.
KW SIGNAL
FT SIGNAL 1 18
FT PROPEP 19 287
FT CHAIN 288 1629
FT DOMAIN 509 587
FT DOMAIN 589 642
FT DOMAIN 645 752
FT DOMAIN 753 880
FT DOMAIN 999 1053
FT DOMAIN 1056 1108
FT DOMAIN 1111 1156
FT DOMAIN 1184 1239
FT DOMAIN 1240 1295
FT DOMAIN 1332 1383
FT DOMAIN 1386 1439
FT DOMAIN 1445 1498
FT DOMAIN 1501 1554
FT DOMAIN 1562 1612
FT DOMAIN 88 96
FT SITE 223 223
FT METAL 434 434
FT ACT_SITE 435 435

*Prediction of the coding sequences of unidentified human genes. XVI.
The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro.*;
DNA Res. 7:65-73(2000).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES. EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG, PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR THYMUS.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration with the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AF261918; AAF89106.1; -.
DR EMBL; AB037733; BAA92550.1; -.
DR HSP; P15167; IATL.
DR MEROPS; M12.021; -.
DR Genew; HGNC:13202; ADAMTS9.
DR MIM; 605421; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_Mtpeptidse.
DR Pfam; PF00090; tsp_1; 11.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 12.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS00092; TSP1; 9.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen; Repeat; Extracellular matrix; Alternative splicing.
KW SIGNAL
FT SIGNAL 1 18
FT PROPEP 19 287
FT CHAIN 288 1629
FT DOMAIN 509 587
FT DOMAIN 589 642
FT DOMAIN 645 752
FT DOMAIN 753 880
FT DOMAIN 999 1053
FT DOMAIN 1056 1108
FT DOMAIN 1111 1156
FT DOMAIN 1184 1239
FT DOMAIN 1240 1295
FT DOMAIN 1332 1383
FT DOMAIN 1386 1439
FT DOMAIN 1445 1498
FT DOMAIN 1501 1554
FT DOMAIN 1562 1612
FT DOMAIN 88 96
FT SITE 223 223
FT METAL 434 434
FT ACT_SITE 435 435

Cysteine switch (POTENTIAL).
Zinc (Catalytic) (By similarity).
By similarity.
```

```
FT METAL 438 438 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 444 444 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 749 749 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 840 840 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1213 1213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1267 1267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1064 1072 CLVTGKGH -> VRWEGCVFP (IN
FT VARSPLIC 1073 1629 MISSING (IN SHORT ISOFORM).
FT CONFLICT 367 367 F -> L (IN REF. 1).
SQ SEQUENCE 1629 AA; 182649 MW; C1C4CEFF58B8941F CRC64;

Query Match 51.2%; Score 1046.5; DB 1; Length 1629;
Best Local Similarity 51.9%; Pred. No. 4.9e-79;
Matches 195; Conservative 57; Mismatches 109; Indels 15; Gaps 8;

QY 1 FVSIPRYVETLVVADESVMKFGADLEHYLLTLTAARLYRHPSILNPINIVVKKVLL 60
DB 288 FLSYPRFVFLVADNRMVSYHGENLQHYILTLMSIVASYIKDPSIGNLNINIVNLVI 347
QY 61 RDRDSGPKVTGNAALTLENFCAWQKLNKVSXKHPYVDYTAITFTRODLCA--TTCDTLG 119
DB 348 HNEQDPSISFNAQITLKNFCQWQHSKNSPGI--HDTAVALLTRQDICRAHDKCDTLG 404
QY 120 MADVTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHDNNKVKVEVFGKLRANHMS 179
DB 405 LAELGTICDPYRSCSISDSGLSTAFTHAGLHGFVNPHDNNKKEE--GVKSPQRVMA 463
QY 180 PTLQIDRANPWSACSAIITDFLDGSHGDCLLDQP--SKPISLPEDLPASYSILSQOCEL 238
DB 464 PTLNFTYTPWMSKSRKYITFEFLDTGYGECLELNEPSPYPLVPQLPGILYNVKNQOCEL 523
QY 239 AFGVGSKPCPYMOYCTKLK---TGKAGQMVQCOTRHFPWADGTSCEGKCLKLGACVER 295
DB 524 IFPGSGQCPYMQCRLWCVNNGVHKG---CRTQHTPWADGTECPGKHCKYGFQCPVK 580
QY 296 HNLNHRVDGSAKWDYPCSRCTGCGGVOLARQCTNPTPANGKYGCEGVRYKRSNCL 355
DB 581 -EMDVPVTDGSGWSWSPETCSRTCGGIGKTAIRECNRPKPKNGGKCYGVGRMKFNSCNT 639
QY 356 EPCPSASGSKSPREQ 371
DB 640 EPCLQKQ--RDRDREQ 653

RESULT 11
AT57_HUMAN STANDARD; PRT; 997 AA.
AC Q9UKP4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7).
GN ADAMTS7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99395124; PubMed=10464288;
RA Hurskainen T.I., Hirohata S., Seidin M.F., Apte S.S.;
RT "ADAM-TS6, ADAM-TS5, and ADAM-TS7, Novel Members of a New Family of
RT Zinc Metalloproteases.";
RL J. Biol. Chem. 274:25555-25563(1999).
CC -|- COPACITOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -|- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG,
```

```
CC LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS.
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -|- PFM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF140675; AAD56358.1; -.
DR HSP; P15167; IATL.
DR MEROPS; M12.231; -.
DR Genew; HGNC:223; ADAMTS7.
DR MIM; 605009; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_Mtpeptidse.
DR Pfam; PF00090; tsp_1; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSP1; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR KW Hydrolyase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 232 BY SIMILARITY.
FT CHAIN 233 997 ADAMTS-7.
FT SITE 204 204 CYSTEINE SWITCH (POTENTIAL).
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 389 389 BY SIMILARITY.
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 462 537 DISINTEGRIN-LIKE.
FT DOMAIN 538 594 TSP TYPE-1 1.
FT DOMAIN 595 697 CYS-RICH.
FT DOMAIN 698 914 SPACER.
FT DOMAIN 915 990 TSP TYPE-1 2.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED02FC104 CRC64;

Query Match 40.18; Score 819; DB 1; Length 997;
Best Local Similarity 47.18; Pred. No. 2.2e-60;
Matches 177; Conservative 43; Mismatches 140; Indels 16; Gaps 10;

QY 2 VSIPIRYVETLVVADESVMKFGH--ADLEHYLLTLTAARLYRHPSILNPINIVVKKVLL 60
DB 238 VSKERKVEVLVADAKVVEYHGPQVSYVLTINMVAGLFHDPSIGNPIHITIVRLVLL 297
QY 61 RDRDSGPKVTGNAALTLENFCAWQKLNKVSXKHPYVDYTAITFTRODLCAAT--CDTL 118
DB 298 EDEEEDLKTTHADNTLTKSPCKWQKSNMKGDAPHLHDDTAILLTRKDLCAAMRNPCEL 357
QY 119 GMADVTGMDPKRSCSVIEDDGLPSAFTAHGLHGFVNPHD--NVKVCVEVFGKLRANHM 177
DB 358 GLSHVAGMCQPHRSCINEDTGLPLATVAHGLHGFQIHDGSGNCEPV--GK--RPF 414
QY 178 MSPTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKP--ISLPEDLPASYSILSQO 236
DB 178 MSPTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKP--ISLPEDLPASYSILSQO 236
```

```
Db 415 MSPQLLYDAAPLTWSCRSHQYITRFLDRGWGLCLDDPPAKDIIIDFPSPVPCVLYDVSHQC 474
QY 237 ELAFGVSGKPCPYM-QYCTKLMCTCKAGQWVCOTRHPFWADGTCGEGKLCCLKGACVER 295
Db 475 RLOYGAYSACFEDMDNVCHTWCSTGVT----TCHSKLDAADVGTGRCGENKWLCSGECVP- 529
QY 296 HNLNKHVRDGNWAKWDPYGPCSRTCGGVQVLARQCTNPTPANGKYGCEGVRYVRYKRCNL 355
Db 530 VGFPRPNDVGGHSGHSANSICSRSCGMGVQSAERQCTOPTPKYKGRVCVGERKFRCLCNL 589
QY 356 EPCPPSASCKSPREQ 371
Db 590 QACP--AGRPSPRHVQ 603

RESULT 12
AT12_HUMAN STANDARD: PRT; 1593 AA.
ID AT12_HUMAN
AC P58397;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-
DE TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=21264577; PubMed=11279086;
RA Cal S., Argueelles J.M., Fernandez P.L., Lopez-Otin C.;
RT "Identification, characterization, and intracellular processing of
RT ADAM-TS12, a novel human disintegrin with a complex structural
RT organization involving multiple thrombospondin-1 repeats.";
RL J. Biol. Chem. 276:17932-17940(2001).
CC 1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC 1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC 1- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC 1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC 1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC 1- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
CC DOMAINS.
CC 1- SIMILARITY: BELONGS TO PERTIDASE FAMILY M12B.
CC 1- SIMILARITY: CONTAINS 8 DISINTEGRIN-LIKE DOMAIN.
CC 1- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ250725; CAC20419.1;
CC Genbank: HGNC:14605; ADAMTS12.
CC MIM: 606184;
CC InterPro: IPR001762; Disintegrin.
CC InterPro: IPR002870; Pep_M12B_propep.
CC InterPro: IPR001590; Reprolysin.
CC InterPro: IPR000884; TSPI.
```

```
DR InterPro: IPR000130; Zn_MTpeptdse.
DR Pfam: PF00090; Tsp_1; 6.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF01562; Pep_M12B_propep; 1.
DR SMART: SM00209; TSPI; 8.
DR PROSITE: PS02015; ADAM_MEPRO; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE: PS00092; TSPI; 2.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 240 BY SIMILARITY.
FT CHAIN 241 1593 ADAMTS-12.
FT DOMAIN 465 544 DISINTEGRIN-LIKE.
FT DOMAIN 545 596 TSP TYPE-1 1.
FT DOMAIN 597 700 CYS-RICH.
FT DOMAIN 701 826 SPACER 1.
FT DOMAIN 827 881 TSP TYPE-1 2.
FT DOMAIN 886 943 TSP TYPE-1 3.
FT DOMAIN 947 995 TSP TYPE-1 4.
FT DOMAIN 996 1315 SPACER 2.
FT DOMAIN 1316 1364 TSP TYPE-1 5.
FT DOMAIN 1367 1423 TSP TYPE-1 6.
FT DOMAIN 1426 1471 TSP TYPE-1 7.
FT DOMAIN 1426 1471 TSP TYPE-1 8.
FT DOMAIN 302 305 POLY-GLU.
FT SITE 208 208 CYSTEINE SWITCH (POTENTIAL).
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 393 393 BY SIMILARITY.
FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 105 105 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 125 125 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 685 685 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 790 790 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 951 951 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1104 1104 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1275 1275 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1300 1300 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1320 1320 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1371 1371 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1378 1378 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1503 1503 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 1593 AA; 177545 MW; 07F9F4863BD83A3 CRC64;

Query Match 38.5%; Score 786; DB 1; Length 1593;
Best Local Similarity 43.1%; Pred. No. 2.1e-57;
Matches 163; Conservative 59; Mismatches 136; Indels 20; Gaps 10;

QY 2 VSIPRYVELLVVADESVMYKEHGD-LEHYLLTLATAARLYRHPISILNPINIVVVKVLL 60
Db 242 ISKERWVELVVDATKMEIYHGSNVEYSILTIMNVTGLFHNPSIGNAHIVVRLILL 301
QY 61 RDRDSGPVKTGNAALTLNFCWAKKLNKVSDEKHPYEDTAIFLTRODLCGA--TTCDTL 118
Db 302 EEEQGLKIVHHAETLSFFCKWQKSNPKSDLPNVHVDVAVLLTRKDCICAGENRPCETL 361
QY 119 GNADVTGKCDPKRCSVLEDDGLPSAFTTAHELGHVFNMPHDNVK-VCSEVFGKLRANHM 177
Db 362 GLSHLSGMCQPHRSRINIEDSGLPLAFTIAHELGHSGIQLHDGKENDCEPVG---RHPYI 418
QY 178 MSPTLIQIDRANPWSACSAIITDFLDSGHGDCILDDQPSKP-ISLPDLPGASYTISQOC 236
Db 419 MSRQLQYDPTPLTWSKCSSEYITRFLDRGWGFCFLDDIPKKKGLSKVIAFGVLDVHHQC 478
QY 237 ELAFGVSGKPCPYM-QYCTKLMCTCKAGQWVCOTRHPFWADGTCGEGKLCCLKGACVER 295
Db 479 QLOYGPNATFCOEVENVCOTLWCSVKG----FCRSKLDAADAAAGTQCGEKKWCMAGKCI-- 532
QY 296 HNLNKHVRDGNWAKWDPYGPCSRTCGGVQVLARQCTNPTPANGKYGCEGVRYVRYKRC 353
```

```

Db 533 -TVGKKPESIFGGMRGWSHSCSRTCGAGVQSAERLCNNPEKFGKYCTGKRYRLC 591
QY 354 NLEPCPSSASGSKSFREQ 371
Db 592 NVHPCRSEA--PTFRQM 607

RESULT 13
AT10_HUMAN
ID AT10_HUMAN STANDARD; PRT; 1077 AA.
AC Q9H324;
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).
GN ADAMTS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Apte S.S.;
RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
RT thrombospondin type 1 repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -|- COPACAPOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -|- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -|- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -|- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF163762; AAG35563.1; .
DR MEPROS; M12.235; .
DR Genew; HGNC:13201; ADAMTS10.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_Mtpeptidse.
DR Pfam; PF00090; tsp.1; 5.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPI; 5.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSPI; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT NON_TER 1
FT PROPEP <1 207 BY SIMILARITY.
FT CHAIN 208 1077 ADAMTS-10.
FT METAL 366 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 367 367 ZINC (CATALYTIC).
FT METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 376 376 ZINC (CATALYTIC).
FT DOMAIN 434 520 DISINTEGRIN-LIKE.
FT DOMAIN 578 679 CYS-RICH.

```

```

FT DOMAIN 680 802 SPACER.
FT FT TSP TYPE-1 1.
FT DOMAIN 521 577
FT DOMAIN 799 860 TSP TYPE-1 2.
FT DOMAIN 862 918 TSP TYPE-1 3.
FT DOMAIN 922 976 TSP TYPE-1 4.
FT DOMAIN 981 1031 TSP TYPE-1 5.
FT CARBOHYD 64 64 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 714 714 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 769 769 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 866 866 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1077 AA; 118072 MW; 3914DE18DCBBF587 CRC64;

Query Match 36.8%; Score 751; DB 1; Length 1077;
Best Local Similarity 41.7%; Pred. No. 1.le-54;
Matches 163; Conservative 55; Mismatches 139; Indels 34; Gaps 11;

QY 2 VSIPIRYVETLVVADESVMKPHG-ADLEHYLLTLTAARLYRHPSILNPINIVVVKVLL 60
Db 209 VSRERYETLVVADKMVMVYHGRDRVEQYVLAINMIVAKLFQDSGLGSTVNILVTRLILL 268
QY 61 RDRDSGPKVTGNAALTNRNFCAMQKLL-----NKVSKRHPYWDTAITLFTQDLG--G 111
Db 269 TEDQPTLEITHAGKSLDSFCKWQKSVNHSGHGNAPENGVAHDTAVLITRYDICIYK 328
QY 112 ATTCDTLCMADVGWCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDNNVKVCEVFG- 170
Db 329 NKPCGTLGLAPVGGMCERSCSYNEDIGLPOAFTIAHEIGTFTGMNHDGVCNSCGARGQ 388
QY 171 ---KLRAHNMSPTLIOIDRANP--WSACAAITDFLDSGHGDCLLDQPSK-PISLPE 224
Db 389 DPAKLMAAHITMKT-----NPFVWSSCRNDYITSLDSGLGLCLNNRPRQDFVYPIV 441
QY 225 LPGASYTLSSQCEALAFGVGSKPCPYMCKWCTGKAKGMVQCRPHFPWADGTCGEG 284
Db 442 APGOAYDAEQCRFQHGKSKQKQKYGVCSELNCLSKSR---CITNSIPAAEGTLCQTH 498
QY 285 KL---CLKGACVERHNLNHRVDGSKWAKWDPYGCPCSTCGGVQVLARRCTNPTPANGG 340
Db 499 TIDKGWCYKRYCVP-FGSRPEGVGANGPWTWPGDCSRTCGGVSSSRHCDSPRTIGG 557
QY 341 KYCEGVRYKYSNLEPCPSSASGSKSFREQ 371
Db 558 KYCLGERRRHRSCTNDDCPGGS--QDFREQ 586

RESULT 14
AT52_HUMAN
ID AT52_HUMAN STANDARD; PRT; 1211 AA.
AC O95450;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 2) (ADAM-TS 2) (ADAM-TS2)
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI)
DE (ADAMTS2 OR PCINP OR PCPNI).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI), AND DISEASE.
RC TISSUE=Skin;
RX MEDLINE=99347935; PubMed=10417273;
RA Colige A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,
RA Wertschke W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,
RA Byers P.H., Lapierre C.M., Prockop D.J., Nusgens B.V.;
RT "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis
are caused by mutations in the procollagen I N-proteinase gene.";

```

Am. J. Hum. Genet. 65:308-317(1999).

-1- FUNCTION: Cleaves the propeptides of type I and II collagen prior to fibril assembly. Does not act on type III collagen. May also play a role in development that is independent of its role in collagen biosynthesis.

-1- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains at Ala-1-Gln.

-1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

-1- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO COLLAGEN TYPE XIV (BY SIMILARITY).

-1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).

-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: LPNPI (SHOWN HERE) AND SPNPI; ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-PROCOLLAGEN PEPTIDASE ACTIVITY.

-1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.

-1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

-1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).

-1- DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos syndrome type VIIC (EDS-VIIC), a recessively inherited connective-tissue disorder characterized clinically by severe skin fragility and joint hypermobility and biochemically by the presence in skin of procollagen incompletely processed at the N-terminus.

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

-1- SIMILARITY: CONTAINS 4 DISINTEGRIN-LIKE DOMAINS.

-1- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AJ003125; CAA05880.1; ..

DR MEROPS: M12.301; ..

DR Genew: HGNC:218; ADAMTS2.

DR MIM: 604539; ..

DR MIM: 225410; ..

DR InterPro: IPR001762; Disintegrin.

DR InterPro: IPR002870; Pep_M12B_propep.

DR InterPro: IPR001590; Reprolysin.

DR InterPro: IPR000884; TSPI.

DR InterPro: IPR000130; Zn_Mtpeptdse.

DR Pfam: PF00090; tsp_1; 4.

DR Pfam: PF01421; Reprolysin; 1.

DR Pfam: PF01562; Pep_M12B_propep; 1.

DR SMART: SM00209; TSPI; 4.

DR PROSITE: PS50215; ADAM_MEPPO; 1.

DR PROSITE: PS50092; TSPI; 1.

DR PROSITE: PS00142; ZINC_PROTEASE; FALSE NEG.

DR PROSITE: PS00427; DISINTEGRIN_1; FALSE NEG.

KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; zymogen;

KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;

Alternative splicing; Ehlers-Danlos syndrome.

SIGNAL 1 29

FT PROPEP 1 BY SIMILARITY.

FT CHAIN 30 253

FT METAL 254 1211

FT ACT_SITE 408 408

FT METAL 409 409

FT METAL 412 412

FT METAL 418 418

FT DOMAIN 480 560

FT DOMAIN 561 617

FT DOMAIN 618 722

FT SITE 691 693

FT SITE 723 851

FT DOMAIN

FT DOMAIN 852 911 TSP TYPE-1 2.

FT DOMAIN 912 974 TSP TYPE-1 3.

FT DOMAIN 975 1030 TSP TYPE-1 4.

FT DOMAIN 40 43 POLY-GLU.

FT DOMAIN 185 188 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 112 112 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 993 993 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1031 1031 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1098 1098 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1145 1145 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1150 1150 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARSPLIC 544 566 HCFGKHCILWLPDILKRDGSMGA -> FRGGAHAACYP

FT VARSPLIC 567 1211 TLGGQGRWIA (IN ISOFORM SPNPI).

FT VARSPLIC 1211 AA; 134722 MW; BECEEF25C3CAD2D CRC64; MISSING (IN ISOFORM SPNPI).

SQ SEQUENCE 1211 AA; 134722 MW; BECEEF25C3CAD2D CRC64; Query Match 31.0%; Score 633.5; DB 1; Length 1211; Best Local Similarity 38.4%; Pred. No. 7.2e-45; Matches 146; Conservative 55; Mismatches 142; Indels 37; Gaps 15;

QY 8 VETLVVADSMVKFHGAD-LEHYLLTLATAARLYRHPSILNPINIVVVKVLLLRDRSG 66

Db 268 IEVLLGVDDSVVDFHGHKEHVQKYLILMNIYVEIHDESLGAHINVLRIILSYGKSM 327

QY 67 PKV-TGNAALTIRNCAWQKLNKVSXKHPEYVDYTAILETRDLCGATTCDTLGMADVGT 125

Db 328 SLIEIGNPSQSLNVCWKWAYLOQKPDTHGVHDHAIETLRQDF-GPSGMQ--GYAPVTG 384

QY 126 MCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDNV-KVCEVEFGKLRANHMMSPTLIQ 184

Db 385 MCFHVSCTLNHEDGFSSAFVVAHETGHLVMEHDCQGNRCGD---EVLGSIAPLVQA 441

QY 185 IDRANPWSACSAIIITDFLDSDGDCLLDQP---SKPISLPEDLPASVTLQQCELAFG 241

Db 442 AFHFHWSRCSQOELSRYLHS--YDCLLDPPAHDWP-ALPQ-LPGLHYSMNEQCRFDFG 497

QY 242 VGSKPDPYMOY---CYKLWCTGKAKGMVQCFHFPAWADTSGCGEGLKLGACV---- 293

Db 498 LGYMMCTAFETFDPCQKLMCS-HPDNPYFCKTKGKPPDLGTMCPAGKHCFCIWLTPD 556

QY 294 --ERHNLNKRVDGWSAKWDPYGCSTRTGGVQLARRQCTNPTPANGKCYCEGVKVKYR 351

Db 557 ILKR-----DGSWGANSPFGSCSTCTGTVKFKTRQCDNPHANGRTCSGLAYDFQ 608

QY 352 SCNLEPCPSASGSKSFREQ 371

Db 609 LCSHQDCPDSLA--DFREQ 626

RESULT 15

ATIS3_HUMAN STANDARD; PRT; 1205 AA.

ID ATIS3_HUMAN

AC O15072; Q9BXZ8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ADAMTS-3 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 3) (ADAM-TS 3) (ADAM-TS3) (Procollagen II (PC

DE amino-proteptide processing enzyme) (Procollagen II N-proteinase) (PC II-NP).

DE II-NP).

GN ADAMTS3 OR KIAA0366.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE OF 1-227 FROM N.A.

EX MEDLINE=21408482;

RA Fernandes R.J., Hirohata S., Engle J.M., Collige A., Cohn D.H.,

RA Eyre D.R., Apte S.S.;

FT *Procollagen II amino propeptide processing by ADAMTS-3. Insights on

dermatoparaxis";
 J. Biol. Chem. 276:31502-31509(2001).
 [2]
 SEQUENCE OF 5-1205 FROM N.A.
 TISSUE=Brain;
 MEDLINE=9734984; PubMed=9205841;
 Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 "Prediction of the coding sequences of unidentified human genes. VII.
 The complete sequences of 100 new cDNA clones from brain which can
 code for large proteins in vitro.";
 DNA Res. 4:141-150(1997).
 CC -!- FUNCTION: Cleaves the propeptides of type II collagen prior to
 fibril assembly. Does not act on types I and III collagens.
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 matrix (by similarity).
 CC -!- TISSUE SPECIFICITY: Found in cartilage and skin.
 CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
 FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
 SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
 CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; AF247668; AAK28400.1; -;
 DR EMBL; AB002364; BAA20821.1; -;
 DR MEROPS; M12.220; -;
 Genew; HGNC:219; ADAMTS3.
 DR MIM; 605011; -;
 DR InterPro; IPR001762; Disintegrin.
 DR InterPro; IPR002870; Pep_M12B_propep.
 DR InterPro; IPR001590; Reprolysin.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR000130; Zn_MTPeptidase.
 DR Pfam; PF000090; tsp_1; 4.
 DR Pfam; PF01421; Reprolysin; 1.
 DR Pfam; PF01562; Pep_M12B_propep; 1.
 DR SMART; SM00209; TSP1; 4.
 DR PROSITE; PS0215; ADAM_MEPRO; 1.
 DR PROSITE; PS50092; TSP1; 2.
 DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
 DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 KW Hydrolyase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
 KW Repeat; Extracellular matrix; Heparin-binding.
 FT SIGNAL 1 20 POTENTIAL.
 FT PROPEP 21 249 BY SIMILARITY.
 FT CHAIN 250 1205 ADAMTS-3.
 FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 399 399 BY SIMILARITY.
 FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 408 408 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 470 550 DISINTEGRIN-LIKE.
 FT DOMAIN 551 607 TSP TYPE-1 1.
 FT DOMAIN 608 712 CYS-RICH.
 FT DOMAIN 713 844 SPACER.
 FT DOMAIN 845 902 TSP TYPE-1 2.
 FT DOMAIN 903 965 TSP TYPE-1 3.
 FT DOMAIN 966 1017 TSP TYPE-1 4.
 FT DOMAIN 246 249 POLY-ARG.
 FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT	CARBOHYD	345	345	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	475	475	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	814	814	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	942	942	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	1205	AA; 135570	MM; EB07B286FC85FB87 CRC64;

Query Match 30.9%; Score 631.5; DB 1; Length 1205;
 Best Local Similarity 38.4%; Pred. No. 1.1e-44;
 Matches 144; Conservative 54; Mismatches 150; Indels 27; Gaps 13;

QY	8	VELTVADDSMYKFGHAD-LEHYLLTLLATAARLYRHPHSILNPINIVVVKVLLLRDRSG	66
Db	258	IEVLGVDDSVVRFGKEHVQNYLLTLMIVNEIYHDESIGVHINVLVRLMILGYAKSI	317
QY	67	PKV-TGNAALTIRNFCAWQKLNKVS DKHPYWDTAIFTRDLCGATTCDTLGMADYGT	125
Db	318	SLIERGNPSRSLNVCWRWASQQORSDLNHEHHDHAIFLTRDFGA---GMQGVAPVTG	374
QY	126	MCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHD---NVKYCEEVFGKLRANHHMSPTL	182
Db	375	MCHPVRSCSTLNHEDGFSSAFVVAHETGHVLMGHDGQGNRCGDETAGMS-----VMAPIV	429
QY	183	IQIDRANPWSACSAAIITDLDGSHGDCILLDOP---SKPISLPEDLPGASYTLSSQCELA	239
Db	430	QRAFHRYHWSRCGGQELKRYIHS--YDCLLDPPFDHWP-KLPE-LPGINYSMDEQCRPD	485
QY	240	FGVSKPCPYMOY---CTKLWCTGKAKGMVQTRHFPWADGTSGEGKLCIKGACVERH	296
Db	486	FGVGYKMCFAFTDFDCKQLWCS-HPDNFYCKTKKGGPLDGTCAAGKWCYKHCMMK-	543
QY	297	NLNKRVDSWAKWDYPGCSRTCGGVOLARRQCTNPTTPANGKYCEGVVRVYKRSCHLE	356
Db	544	NANQOKQDGNWGSWTKFGSCSRTCGTGVFRTRQCNPNPMPINGGQDCPGVNFYQLCNTE	603
QY	357	PCPSSASGSKSREEQ	371
Db	604	ECQKHF--EDFRAQQ	616

Search completed: April 29, 2003, 17:14:00
 Job time : 9.56924 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:25 ; Search time 33.4317 Seconds
(without alignments)
2738.218 Million cell updates/sec

Title: US-10-009-332-1-copy_1_687

Perfect score: 3744

Sequence: 1 MLLLGILTLAFAGTAGGFE.....RFDKCGVCGDNKCKKVTG 687

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3744	100.0	950	22	Human metalloprotease
2	3736	99.8	950	23	Human protease #2.
3	3725.5	99.5	928	23	Human metalloprotease
4	3725.5	99.5	952	23	Human protease PRT
5	2497.5	66.7	505	21	Rat metalloprotease
6	1966	52.5	381	21	Human metalloprotease
7	1960.5	52.4	950	20	Human METH1 protei
8	1960.5	52.4	950	22	Human ADAM-type me
9	1960.5	52.4	950	22	Human METH1. Homo
10	1960.5	52.4	967	19	Human integrin lig

11	1960.5	52.4	968	22	AA550011	Protein; SEQ ID 12
12	1959.5	52.3	967	20	AA504142	Human Tango-71 pro
13	1947.5	52.0	896	21	AA521265	Mouse metalloprote
14	1938.5	51.8	967	20	AA578189	Human secreted pro
15	1921	51.3	950	21	AA53899	Amino acid sequenc
16	1913	51.1	367	23	AA522542	Human protease #3.
17	1751	46.8	727	20	AA578435	Human ADAMTS-1 pro
18	1699.5	45.4	890	20	AA549502	Human METH2. Homo
19	1699.5	45.4	890	22	AA550003	Human ADAM type me
20	1696.5	45.3	889	22	AA574946	Human aggrecan deg
21	1689.5	45.1	837	20	AA575425	Human ADAM type me
22	1688.5	45.1	837	21	AA599429	Human aggrecanase
23	1688.5	45.1	837	22	AA521919	Human PRO1563 (UNQ
24	1688.5	45.1	837	22	AA566178	Protein of the inv
25	1687.5	45.1	837	22	AA578228	Human aggrecanase
26	1687.5	45.1	840	21	AA521256	Human metalloprote
27	1679	44.8	905	22	AA572284	Murine ADAMTS-8 am
28	1669	44.6	321	23	AA522540	Human protease #1.
29	1611	43.0	930	22	AA572280	Murine ADAMTS-5 am
30	1588.5	42.4	870	21	AA521252	Rat metalloproteai
31	1587.5	42.4	930	20	AA575426	Human aggrecan deg
32	1580	42.2	929	21	AA541226	Human ORFX ORF990
33	1543.5	41.2	1629	23	AA530703	Human aggrecanase
34	1543.5	41.2	1629	23	AA514448	Human ADAMTS-SI pr
35	1543.5	41.2	1916	23	AA519173	Human protease, PR
36	1543.5	41.2	1935	23	AA572896	Human metalloprote
37	1540.5	41.1	1602	23	AA530702	Human aggrecanase
38	1539.5	41.1	947	22	AA586950	Human metalloprote
39	1537.5	41.1	1073	21	AA521264	Human metalloprote
40	1523	40.7	1934	22	AA572301	Human ADAMTS-9 alt
41	1503.5	40.2	625	23	AA548394	Rat aggrecanase.
42	1472	39.3	1505	23	AA572897	Human metalloprote
43	1466	39.2	680	21	AA521251	Human metalloprote
44	1460.5	39.0	1907	23	AA571133	Human protease #12
45	1432	38.2	1882	22	AA572286	Human ADAMTS-9 ami

ALIGNMENTS

RESULT 1
AAG52299
ID AAG62299 standard; protein; 950 AA.
XX
AC AAG62299;
XX
DT 23-AUG-2001 (first entry)
XX
DE Human metalloprotease MDT56 protein.
XX
KW Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;
KW osteopathic; antiarthritic.
XX
OS Homo sapiens.
XX
PN WO200134785-A1.
XX
PD 17-MAY-2001.
XX
PF 10-NOV-2000; 2000WO-JP07917.
XX
PR 11-NOV-1999; 99JP-0321740.
PR 16-MAY-2000; 2000JP-0144020.
XX
PA (YAMA) YAMANOUCHI PHARM CO LTD.
PA (KAZU-) KAZUSA DNA RES INST.
PI Yamaaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;
XX WPI: 2001-343602/36.
XX N-PSDB: AAH41003.
PT Metalloprotease with aggrecanase activity for treating joint diseases

PT especially osteoarthritis -
XX
PS Claim 1; Page 56-60; 85pp; Japanese.
XX
CC This invention relates to a metalloprotease with aggreganase activity.
CC The invention includes protein and DNA sequences of the metalloprotease,
CC vectors containing the DNA, host cells transformed by the vectors, and
CC antibodies directed against the metalloprotease. The antibodies, protein
CC and DNA sequences can be used in the treatment and prevention of joint
CC diseases, particularly osteoarthritis. The treatment may result in
CC osteopathic and antiarthritic activity. The present sequence represents
XX the metalloprotease of the invention termed MDTs6.
XX
SQ Sequence 950 AA;

Query Match 100.0%; Score 3744; DB 22; Length 950;
Best Local Similarity 100.0%; Pred. No. 4.4e-285;
Matches 687; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLGILTLAFAGRTAGGFEPEREVVPVIRLDPDINGRRYYWRGPEDSGDQGLIFQITAF 60
DB 1 MLLGILTLAFAGRTAGGFEPEREVVPVIRLDPDINGRRYYWRGPEDSGDQGLIFQITAF 60

QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDPSFAAVSLC 120
DB 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDPSFAAVSLC 120

QY 121 GGLRGAFGYRGAEYVISPPLNASAPAAQNSOGAHLQRRGVPGSGDPTSRGCVASGW 180
DB 121 GGLRGAFGYRGAEYVISPPLNASAPAAQNSOGAHLQRRGVPGSGDPTSRGCVASGW 180

QY 181 NPAILRALDPYKRRAGFGESESRSSRRSGRAKRFVSIPIRYVETLVVADESMVKFHGADLEH 240
DB 181 NPAILRALDPYKRRAGFGESESRSSRRSGRAKRFVSIPIRYVETLVVADESMVKFHGADLEH 240

QY 241 YLLTLATAARLYRHPSILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWQKKLN 300
DB 241 YLLTLATAARLYRHPSILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWQKKLN 300

QY 301 KVS DKHPEYWDTAILFTRODLGATTCDTLGMADVGTMC DPRKSCSVIEDDGLPSAFTTA 360
DB 301 KVS DKHPEYWDTAILFTRODLGATTCDTLGMADVGTMC DPRKSCSVIEDDGLPSAFTTA 360

QY 361 HELGHVFNMPHDNVKVEEVFGKLRANHMSP TLIQIDRANPWSACSAIITDFLDSHG 420
DB 361 HELGHVFNMPHDNVKVEEVFGKLRANHMSP TLIQIDRANPWSACSAIITDFLDSHG 420

QY 421 DCILLDQSPKPISLPEDLPFGASYTLSQOCELAFCVGSKPCPYMOYCTKLWCTGKAKGQMYC 480
DB 421 DCILLDQSPKPISLPEDLPFGASYTLSQOCELAFCVGSKPCPYMOYCTKLWCTGKAKGQMYC 480

QY 481 QTRHFPWADGTSCEGKLCILKACVERHNLNKHVRVDSWAKMDPYGPCSRTCGGVQLAR 540
DB 481 QTRHFPWADGTSCEGKLCILKACVERHNLNKHVRVDSWAKMDPYGPCSRTCGGVQLAR 540

QY 541 RQCTNPTPANGKYCEGVRYKRSCLNLEPCPSASGKSPREOCEAFNGYNHSTNLTIA 600
DB 541 RQCTNPTPANGKYCEGVRYKRSCLNLEPCPSASGKSPREOCEAFNGYNHSTNLTIA 600

QY 601 VAWPKYSGVSPRDKKLCICRANGTGYFYVLPKVVDTGLCSPDSTVVCVQGCIRKAGCD 660
DB 601 VAWPKYSGVSPRDKKLCICRANGTGYFYVLPKVVDTGLCSPDSTVVCVQGCIRKAGCD 660

QY 661 GNLSGKRRFDCGVCVGGDNKSKKVTG 687
DB 661 GNLSGKRRFDCGVCVGGDNKSKKVTG 687

RESULT 2
AAE22541
ID AAE22541 standard; Protein; 950 AA.
XX
AC AAE22541;

XX 26-JUL-2002 (first entry)
DT
XX Human protease #2.
DE
XX Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme.
XX
OS Homo sapiens.
XX
FN WO200226949-A2.
XX
PD 04-APR-2002.
XX
PF 27-SEP-2001; 2001WO-US30350.
XX
PR 29-SEP-2000; 2000US-236689P.
XX
PA (LEXI-) LEXICON GENETICS INC.
PI Friddle CJ, Hilbun E;
XX
DR WPI; 2002-372123/40.
DR N-PSDB; AAD35569.
XX
PT Novel nucleic acid encoding a human protease, useful as a hybridization
PT probe for screening libraries and assessing gene expression patterns -
XX
PS Claim 6; Page 36-38; 41pp; English.
XX
CC The present sequence is novel human protein (NHP), human protease.
CC NHPs share structural similarity with animal proteases particularly
CC zinc metalloproteases. Sequences of the invention are useful in
CC therapeutic, diagnostic and pharmacogenomic applications. NHP
CC polynucleotides are used as hybridization probes for screening
CC libraries and assessing gene expression patterns. They can also be
CC used for treating related biological disorders such as obesity, high
CC blood pressure, arthritis, connective tissue disorders and infertility.
XX They are also used in gene therapy.
SQ Sequence 950 AA;

Query Match 99.8%; Score 3736; DB 23; Length 950;
Best Local Similarity 99.9%; Pred. No. 1.9e-284;
Matches 686; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLLGILTLAFAGRTAGGFEPEREVVPVIRLDPDINGRRYYWRGPEDSGDQGLIFQITAF 60
DB 1 MLLGILTLAFAGRTAGGFEPEREVVPVIRLDPDINGRRYYWRGPEDSGDQGLIFQITAF 60

QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDPSFAAVSLC 120
DB 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFYSGDVNAEPDPSFAAVSLC 120

QY 121 GGLRGAFGYRGAEYVISPPLNASAPAAQNSOGAHLQRRGVPGSGDPTSRGCVASGW 180
DB 121 GGLRGAFGYRGAEYVISPPLNASAPAAQNSOGAHLQRRGVPGSGDPTSRGCVASGW 180

QY 181 NPAILRALDPYKRRAGFGESESRSSRRSGRAKRFVSIPIRYVETLVVADESMVKFHGADLEH 240
DB 181 NPAILRALDPYKRRAGFGESESRSSRRSGRAKRFVSIPIRYVETLVVADESMVKFHGADLEH 240

QY 241 YLLTLATAARLYRHPSILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWQKKLN 300
DB 241 YLLTLATAARLYRHPSILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAWQKKLN 300

QY 301 KVS DKHPEYWDTAILFTRODLGATTCDTLGMADVGTMC DPRKSCSVIEDDGLPSAFTTA 360
DB 301 KVS DKHPEYWDTAILFTRODLGATTCDTLGMADVGTMC DPRKSCSVIEDDGLPSAFTTA 360

QY 361 HELGHVFNMPHDNVKVEEVFGKLRANHMSP TLIQIDRANPWSACSAIITDFLDSHG 420
DB 361 HELGHVFNMPHDNVKVEEVFGKLRANHMSP TLIQIDRANPWSACSAIITDFLDSHG 420

Db 361 HELGHVFNPHDNVNVKCEEVFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSGHG 420
 QY 421 DCLLDQPSKPISLPEDLPASVYTLSSQCELAFCVGSKPCPYMOYCTKLWCTGKAKGQMYC 480
 Db 421 DCLLDQPSKPISLPEDLPASVYTLSSQCELAFCVGSKPCPYMOYCTKLWCTGKAKGQMYC 480
 QY 481 QTRHFPWADGTSCEGKCLKLGACVERHNLNKHRVDGWSAKWDPYPCSRCTCGGGVQLAR 540
 Db 481 QTRHFPWADGTSCEGKCLKLGACVERHNLNKHRVDGWSAKWDPYPCSRCTCGGGVQLAR 540
 QY 541 ROCTNPTPANGKYCEGVVRKYRSCNLEPCPSSASGKSFREQCCEAFNGYNHSTNRLTLA 600
 Db 541 ROCTNPTPANGKYCEGVVRKYRSCNLEPCPSSASGKSFREQCCEAFNGYNHSTNRLTLA 600
 QY 601 VAWVPKYSVSPRDKCKLICRANGTGYFYVLAPKVVDGTLCSPDSTSVCVQKCKIKAGCD 660
 Db 601 VAWVPKYSVSPRDKCKLICRANGTGYFYVLAPKVVDGTLCSPDSTSVCVQKCKIKAGCD 660
 QY 661 GNLGSKKRFDKCGVCGGDNKSKCKVTG 687
 Db 661 GNLGSKKRFDKCGVCGGDNKSKCKVTG 687
 RESULT 3
 AAU72899
 ID AAU72899 standard; Protein: 928 AA.
 AC AAU72899;
 DT 26-FEB-2002 (first entry)
 XX Human metalloprotease partial protein sequence #11.
 KW Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;
 KW vasotropic; antimigraine; analgesic; endocrine; neurotropic; tranquiliser;
 KW hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;
 KW anorectic; antiinflammatory; aspartyl protease; cysteine protease;
 KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;
 KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;
 KW immune-related disease; cardiovascular disease; neuronal disease;
 KW migraine; sexual dysfunction; mood disorder; attention disorder;
 KW cognition disorder; hypotension; hypertension; psychotic disorder;
 KW dyskinesia; metabolic disorder; inflammatory disorder.
 OS Homo sapiens.
 XX
 PN WO200183782-A2.
 XX
 PD 08-NOV-2001.
 XX
 PF 04-MAY-2001; 2001WO-US14431.
 XX
 PR 04-MAY-2000; 2000US-201879p.
 XX
 PA (SUGEN-) SUGEN INC.
 PI Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
 PI Payne V;
 XX
 DR WPI; 2002-041502/05.
 DR N-PSDB; AAS97182.
 XX
 PT Novel protease polypeptide useful for screening for substances that may
 PT be used to treat, e.g., cancers, immune-related diseases,
 PT cardiovascular disease, migraine, pain, psychotic and inflammatory
 PT disorders -
 XX
 PS Claim 28; Figure 2G; 232pp; English.
 XX
 CC The invention relates to an isolated, enriched, or purified protease
 CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
 CC screen for substances (S) that may modulate its activity. Administering
 CC S (which modulates protease activity in vitro) may be used to treat a

CC disease or disorder selected from cancers (e.g., of tissues, of blood or
 CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,
 CC brain, ovarian, bladder or kidney), immune-related diseases and
 CC disorders, cardiovascular disease, brain or neuronal-associated diseases
 CC (e.g., central or peripheral nervous system diseases, migraine, pain,
 CC sexual dysfunction, mood disorders, attention disorders, cognition
 CC disorders, hypotension, hypertension, psychotic disorders, neurological
 CC disorders and dyskinesias), metabolic disorders and inflammatory
 CC disorders. (I) may also be useful as a diagnostic tool for a disease or
 CC disorder such as those above. AAU72876-AAU72910 represent human
 CC protease amino acid sequences of the invention.
 XX
 SQ Sequence 928 AA;

Query Match 99.5%; Score 3725.5; DB 23; Length 928;
 Best Local Similarity 99.7%; Pred. No. 1.2e-283;
 Matches 686; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 1 MLLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPDGQGLIFQITAF 60
 Db 26 MLLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPDGQGLIFQITAF 85
 QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRCFCYSGDVNAEPDFAAVSLC 120
 Db 86 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRCFCYSGDVNAEPDFAAVSLC 145
 QY 121 GGLRGAEGYRGAEVVISPLNAPAAQNSQGAHLQRRGVPGGSGDPTSRGCVASGW 180
 Db 146 GGLRGAEGYRGAEVVISPLNAPAAQNSQGAHLQRRGVPGGSGDPTSRGCVASGW 205
 QY 181 NPAILRALDPYKPRRAGFESRRRRSGRAKRFVSPRYVETLVVADESVMKFGADLEH 240
 Db 206 NPAILRALDPYKPRRAGFESRRRRSGRAKRFVSPRYVETLVVADESVMKFGADLEH 265
 QY 241 YLTLTAARLYRHPHSILNPINIVVYVLLLRDRDGGPKVTGNAALTFRNFCAWQKLN 300
 Db 266 YLTLTAARLYRHPHSILNPINIVVYVLLLRDRDGGPKVTGNAALTFRNFCAWQKLN 325
 QY 301 KVSDDKHPEYWDTAILFRQDLCCATTCTDLGMADVTGMDPKSCSVIEDDGLPSAFTTA 360
 Db 326 KVSDDKHPEYWDTAILFRQDLCCATTCTDLGMADVTGMDPKSCSVIEDDGLPSAFTTA 385
 QY 361 HELGHVFNPHDNVNVKCEEVFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSGHG 420
 Db 386 HELGHVFNPHDNVNVKCEEVFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSGHG 445
 QY 421 DCLLDQPSKPISLPEDLPASVYTLSSQCELAFCVGSKPCPYMOYCTKLWCTGKAKGQMYC 480
 Db 446 DCLLDQPSKPISLPEDLPASVYTLSSQCELAFCVGSKPCPYMOYCTKLWCTGKAKGQMYC 505
 QY 481 QTRHFPWADGTSCEGKCLKLGACVERHNLNKHRVDGWSAKWDPYPCSRCTCGGGVQLAR 540
 Db 506 QTRHFPWADGTSCEGKCLKLGACVERHNLNKHRVDGWSAKWDPYPCSRCTCGGGVQLAR 565
 QY 541 ROCTNPTPANGKYCEGVVRKYRSCNLEPCPSSASGKSFREQCCEAFNGYNHSTNRLTLA 600
 Db 566 ROCTNPTPANGKYCEGVVRKYRSCNLEPCPSSASGKSFREQCCEAFNGYNHSTNRLTLA 625
 QY 601 VAWVPKYSVSPRDKCKLICRANGTGYFYVLAPKVVDGTLCSPDSTSVCVQKCKIKAGC 659
 Db 626 VAWVPKYSVSPRDKCKLICRANGTGYFYVLAPKVVDGTLCSPDSTSVCVQKCKIKAGC 685
 Db 686 DGNLGSKKRFDKCGVCGGDNKSKCKVTG 713

RESULT 4
 AAU74751
 ID AAU74751 standard; Protein: 952 AA.
 XX
 AC AAU74751;
 XX

109-APR-2002 (first entry)

Human protease PRPS-11 protein sequence.

Human; protease; PRPS; gastrointestinal; Crohn's disease; cancer; cardiovascular; atherosclerosis; autoimmune deficiency; dermatitis; inflammatory disorder; acquired immunodeficiency syndrome; AIDS; cell proliferative disorder; developmental disorder; epilepsy; Duchenne muscular dystrophy; epithelial disorder; neurological disorder; reproductive disorder; endometriosis.

Homo sapiens.

WO200198468-A2.

27-DEC-2001.

13-JUN-2001; 2001WO-US19178.

16-JUN-2000; 2000US-212336P.

22-JUN-2000; 2000US-213955P.

29-JUN-2000; 2000US-215396P.

07-JUL-2000; 2000US-216821P.

14-JUL-2000; 2000US-218946P.

(INCY-) INCYTE GENOMICS INC.

Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM; Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA; Wallia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT; Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L; Kallack DA;

WPI: 2002-090437/12.

N-PSDB; ABK12894.

Twenty one human proteases (referred to as PRPS-1 to PRPS-21), useful in the diagnosis, treatment and prevention of gastrointestinal (e.g. gastritis), cardiovascular (e.g. atherosclerosis) and cell proliferative (e.g. cancer) disorders -

Claim 1; Page 144-146; 177pp; English.

The present invention relates to twenty one new human proteases, referred to as PRPS-1 to PRPS-21. The PRPS polynucleotides and polypeptides of the invention are useful in the diagnosis, treatment and prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and myocardial infarction, autoimmune/inflammatory e.g. acquired immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell proliferative e.g. cancer, developmental e.g. Duchenne and Becker muscular dystrophy, epithelial e.g. dermatitis, neurological e.g. epilepsy and Alzheimer's disease and reproductive e.g. infertility and endometriosis disorders. Numerous other examples of each disorder are given in the specification. The present protein sequence represents the human protease PRPS-11 protein of the invention.

Sequence 952 AA;

Query Match 99.5%; Score 3725.5; DB 23; Length 952;

Best Local Similarity 99.7%; Pred. No. 1.2e-283;

Matches 686; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 MLLGILFLAFAGTAGGPEREVVPIRLDPDINGRRYWRGPDSDGGLIFQITAF 60

DB 1 MLLGILFLAFAGTAGGPEREVVPIRLDPDINGRRYWRGPDSDGGLIFQITAF 60

QY 61 QEDFYLHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFYSGDVNAEPDSFAAUSLC 120

DB 61 QEDFYLHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFYSGDVNAEPDSFAAUSLC 120

QY 121 GGLRGAFGYRGAEEYVISPLPNASAPAAQRNSQGAHLLQRRGVPGGSDPTSRGCVASGW 180

Db 121 GGLRGAFGYRGAEEYVISPLPNASAPAAQRNSQGAHLLQRRGVPGGSDPTSRGCVASGW 180

QY 181 NPATILRALDPYKPRRAGFSGESRRRRSGRAKRFVSIIPRYVETLWVADESVMVKFHGADLEH 240

Db 181 NPATILRALDPYKPRRAGFSGESRRRRSGRAKRFVSIIPRYVETLWVADESVMVKFHGADLEH 240

QY 241 YLLTLLATAARLYRHPSTLNPINIVVVKVLLLRDRSDGPKVTGNAALTNRNFCAWOKKLN 300

Db 241 YLLTLLATAARLYRHPSTLNPINIVVVKVLLLRDRSDGPKVTGNAALTNRNFCAWOKKLN 300

QY 301 KVSOKHPEYWDATILFTRODLGATTCDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTA 360

Db 301 KVSOKHPEYWDATILFTRODLGATTCDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTA 360

QY 361 HELGHVFNMHPDHNKVCVEEFGKLRANHMSPPTLIQIDRANPWSACSAAIITDPLDSGHG 420

Db 361 HELGHVFNMHPDHNKVCVEEFGKLRANHMSPPTLIQIDRANPWSACSAAIITDPLDSGHG 420

QY 421 DCLLDQPSKPISLPEDLPASYSYTLSSQCELAFGVSGKPCPYMXYCTKLWCTGKAKGQWVC 480

Db 421 DCLLDQPSKPISLPEDLPASYSYTLSSQCELAFGVSGKPCPYMXYCTKLWCTGKAKGQWVC 480

QY 481 QTRHFPWADGTSCGEGKLCGLKAGACVERHNLNKHRVDGSKWAKWDPYGPCSRCTCGGVQLAR 540

Db 481 QTRHFPWADGTSCGEGKLCGLKAGACVERHNLNKHRVDGSKWAKWDPYGPCSRCTCGGVQLAR 540

QY 541 ROCTNPTFANGKYCEGVRYKRSNLEPCPSSASGKSFREEQCEAFNGYNHSTNRLTLA 600

Db 541 ROCTNPTFANGKYCEGVRYKRSNLEPCPSSASGKSFREEQCEAFNGYNHSTNRLTLA 600

QY 601 VAWVPKYSVSPRDKCKLICRANGTYFYVLAPK-VVDGTLCSPDSTSVQVQKCIKAGC 659

Db 601 VAWVPKYSVSPRDKCKLICRANGTYFYVLAPK-VVDGTLCSPDSTSVQVQKCIKAGC 660

QY 660 DGNLGSKKRFRDKCGVCGGDNKSKKVTG 687

Db 661 DGNLGSKKRFRDKCGVCGGDNKSKKVTG 688

RESULT 5

AAB21257

ID AAB21257 standard; Protein; 505 AA.

XX

AC AAB21257;

XX

DT 23-FEB-2001 (first entry)

XX

DE Rat metalloproteinase ADAMTS-5.

XX

KW Rat; ADAMTS-5; metalloproteinase; ADAM;

KW a disintegrin and metalloproteinase domain; thrombospondin domain;

KW vaccine; neutropic; neuroprotective; antiparkinsonian;

KW cerebroprotective; cytosolic; antiarthritic; immunosuppressive;

KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;

KW autoimmune disease; brain tumour; brain injury.

XX

OS Rattus norvegicus.

XX

PN WO2000053774-A2.

XX

PD 14-SEP-2000.

XX

PF 08-MAR-2000; 2000WO-US06237.

XX

PR 08-MAR-1999; 99US-0264585.

XX

PA (NEUR-) NEUROCRINE BIOSCIENCES INC.

XX

PI Kelner GS, Clark M, Maki RA;

XX

DR WPI: 2000-594326/56.

DR N-PSDB; AAA95827.

XX

KW a disintegrin and metalloproteinase domain; thrombospondin domain;
KW vaccine; neurotropic; neuroprotective; antiparkinsonian;
KW cerebroprotective; cytostatic; antiarthritic; immunosuppressive;
KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
KW autoimmune disease; brain tumour; brain injury.

	Location/Qualifiers
Key	196
Misc-difference	/note= "Xaa= any amino acid"
FT	268
Misc-difference	/note= "Xaa= any amino acid"
FT	304
Misc-difference	/note= "Xaa= any amino acid"
FT	308
Misc-difference	/note= "Xaa= any amino acid"
FT	
FT	
XX	

PN WO200053774-A2.
XX
XX
PD 14-SEP-2000.
XX
XX
PF 08-MAR-2000; 2000WO-US06237.
XX

PR 08-MAR-1999; 9905-0264585.
XX

XX
XX
XX

(NEW) NEUROKINE BIOSCIENCES INC.

XX
 00
 WDT: 2000-504206/56

DR N-PSDB; AAA95831.
XX

PT Polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to

XX
XX
XX

prevent and treat Alzheimer's disease, cancer and autoimmune diseases

[illegible]

CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin and Metalloprotease) family. Members of the ADAMTS family

CC contain a thrombospondin domain in addition to the disintegrin and
CC metalloproteinase domains found in the ADAMTS polypeptides are

CC useful for the manufacture of medicaments for treating conditions
CC associated with neuroinflammation and/or neurodegeneration, such as

CC useful for treating conditions associated with cell proliferation, cell migration, inflammation and/or angiogenesis such as cancer, arthritis

CC and autoimmune diseases. They can be used to treat patients afflicted with an invasive tumour, a brain tumour or brain injury.

XX	SQ	Sequence	381 AA;
----	----	----------	---------

Query Match 52.5%; Score 1966; DB 21; Length 381;

Matches 358; Conservative 3; Mismatches 18; Indels 2; Gaps 2

Qy 240 HYLTLATAARLYRHRHSILNPINIVVKVLLLRDSDGPKVTGNAALTRLNFCWQKKL 299

DB 2 HYRARRAARAG-IFKHPSILNPIINVVKVLLLRDRDSGPKVTGNAALTLRNFCAWQKKL 60

QY 300 NNSDKHFEINDIALEFTRQDLECGAICDILGTADAVIMCDFKRSQSVLEDDGLFSAT11 333
|||||
34 |||||

350. AINDE CHUENNAHUNNAKIVETEVETCI DANUMMOCMI TOTODANDWACACNA ITIMEDI DOCH 410

121 AHEIGHVENMPBDNVKCFEEVEGKIRANHHMSPITLIDRANPWSACSAATTORIDSGH 180

QY 420 GDCLLDQSPKPSISLPEDIPGASYTLSQQCELAFGVGSKPCPYMQYCTKLWCTGKAKGMV 479

Db 181 GDCLDQPSKPIFLPXDLPGASYTLSQQCELAFGVGFKPCPYMQYCTKLWCTGKAKGMV 240

QY 480 CQTRHPWADGTCSEGGKCLKAGACVERHNLNKRVDSWAKWDPYGCSCRTCGGGVOLA 539
 DB 241 CQTRHPWADGTCSEGGKCLKAGACVEXHNLNKRVDWAKWDPYGCSCRTCGGGVOLA 300
 QY 540 RQCTNPTP-ANGGKYCGVRKYRSCNLEPCPSASGKSPREOCFAFNGYNHSTNRLT 598
 DB 301 RQXHOXPPLPGGKYCGVRKYRSCNLEPCPSASGKSPREOCFAFNGYNHSTNRLT 360
 QY 599 LAVAWPKYSGVSPRDKCKLI 619
 DB 361 LAVAWPKYSGVSPRDKCKLI 381

RESULT 7
 AAY49501
 ID AAY49501 standard; Protein; 950 AA.
 XX
 AC AAY49501;
 DT 10-JAN-2000 (first entry)
 DE Human METH1 protein.
 KW Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;
 KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
 KW angiogenesis inhibitor; abnormal wound healing; inflammation;
 KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
 KW diabetic retinopathy; macula degeneration; haemangioma; detection;
 KW arterial-venous malformation; immune deficiency.
 XX
 OS Homo sapiens.
 XX
 PN WO9937660-A1.
 XX
 PD 29-JUL-1999.
 XX
 PF 22-JAN-1999; 99WO-US01313.
 XX
 PR 23-JAN-1998; 98US-0072298.
 PR 28-AUG-1996; 98US-0098539.
 XX
 PA (IRUELA) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 XX
 PI Iruela-Arispe L, Hastings GA, Ruben SM;
 XX
 DR WPI; 1999-590684/50.
 DR N-PSDB; AA232000.
 XX
 PT New isolated metalloprotease thrombospondin polypeptides, useful for
 PT treating hyperproliferative disorders, cancers or autoimmune disorders
 XX
 PS Claim 10; Fig 1; 457pp; English.

AA232000 and AA232001 encode, and AAY49501 and AAY49502 represent, human
 metalloprotease thrombospondin (METH) proteins METH1 and METH2
 respectively. METH1 and METH2 have been found to be potent inhibitors of
 angiogenesis both in vitro and in vivo. They can be used for treating
 cancer and other disorders related to angiogenesis including abnormal
 wound healing, inflammation, rheumatoid arthritis, psoriasis,
 endometrial bleeding disorders, diabetic retinopathy, some forms of
 macula degeneration, haemangiomas, and arterial-venous malformations.
 They may be useful in treating deficiencies or disorders of the immune
 system, by activating or inhibiting the proliferation, differentiation,
 or mobilisation (chemotaxis) of immune cells. The etiology of these
 immune deficiencies or disorders may be genetic, somatic, such as
 cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
 toxins), or infectious. They can also be used to treat inflammatory or
 conditions, both chronic and acute conditions. The products can also be
 used for detection and diagnosis. AA232002 to AA232080, and AAY49503 to

CC AAY49511 represent sequences given in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 950 AA;
 Query Match 52.4%; Score 1960.5; DB 20; Length 950;
 Best Local Similarity 53.6%; Pred. No. 4.7e-145;
 Matches 388; Conservative 91; Mismatches 178; Indels 67; Gaps 17;
 QY 1 MLLLGILTLAFAGRTAGGFEPEVVRVPIRLDPDINGRRYRWGPEDSGQGLIFQITAF 60
 DB 19 LLLLAALLAVSDALGRPSEDEELVVP-ELE-----RAP---GHGTTLRLLHAF 64
 QY 61 QEDFVHLHTPDPAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRCFYSGDVNAEPDSF 114
 DB 65 DQDLLELRPDSSFLAFOFTLQNGRKSGSETPLP-----ETDLAHCFYSGTVNGDPSSA 119
 QY 115 AAVSLCGGLRGAFGYRGAEYVISPLPNAS-----APAAQRNSQGA-----HLIQ--RRGVPGG 165
 DB 120 AALSCEGVGAFYLLGEAYFIQPLPAASERLATAAPGEKPPAPLQPHLLRRNRQDVG 179
 QY 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPYKPRRAGFSGSRRRS 207
 DB 180 TCGVVDDPRPTGKAETEDGEGEGPQWS-----PQDPALQGVGP-TGTGS 230
 QY 208 GRAKFEVPIRYVEFLVVADESVMYKFGADLEHYLLTLLATAARLYRHPISILNPINIVV 267
 DB 231 IRKRFVSHRYVETMLVADQSMAEFGHGLKHYLLLFVSAARLYKHPISIRNSVSLVV 290
 QY 268 KYLLLRDRSGPKVTGNAAALTRNFCWOKKLNKVSFKHPEYWDATLFTRDQLCGATTC 327
 DB 291 KILVHDEQKGEVTSNAAALTRNFCWOKHNPSPDRDAEHYDTALLFTRDQLCGSQT 350
 QY 328 DTLGMADVTMCDPKRSCTVEDDGLPSAFTTAHELGHVFMNPHNDVNVKVEEFGKLRAN 387
 DB 351 DTLGMADVTCTDPSRSCSVIEDDGLQAAFTTAHELGHVFMNPHDDAKQACASLNGVQDS 410
 QY 388 HWSPTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPSLSPEDLPASLYTL 447
 DB 411 HMASMLNLDHSQWSPSCSAYMITSFLDNGHGECLEMDKFNQPIQLPGDLPSTYDANRQ 470
 QY 448 CELAFGVGSKPCP-YMOYCTKLWCTGKAKGQWCVOTRHFPWADGTCSEGGKCLKAGC 506
 DB 471 CQTFEGEDSKHCPDAASTCTLWCTGTSGGVVLCQTHFPWADGTCSEGGKWCINGKCVN 530
 QY 507 RHNLNKH--RVDSWAKWDPYGCSCRTCGGGVOLARROCTNPTPANGKYCYGVRVYR 563
 DB 531 KTD-RKHEDTFPHGSGWGMGPGWDCSTTCGGVQYTHRECDNVPVNGGKYCEGKRVYR 589
 QY 564 SCNLEPCPSSASGKSFREEQCAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRAN 623
 DB 590 SCNLEDCPDN-NGKTFREEQCAHNEFSKASFGSGPAVEWIPKYAGVSPKDRCKLICQAK 648
 QY 624 GGYEYVLAPKVVDGTLCSPDSTSVCGQKIKAGCDNGLSKKRDPKCCVCGDNKSK 683
 DB 649 GIGYEFYVLPQKVVDGTPCSPDSTSVCGQCVKAGCDRIIDSKKFKDKCGVCGGNGSTCK 708
 QY 684 KVTG 687
 DB 709 KISG 712
 RESULT 8
 AAB73549
 ID AAB73549 standard; Protein; 950 AA.
 XX
 AC AAB73549;
 DT 07-AUG-2001 (first entry)
 XX
 DE Human ADAM-type metalloprotease MDT54, SEQ ID NO:4.
 XX
 KW Human; MDT54; ADAM-type metalloprotease; drug screening;

PT rheumatoid arthritis and psoriasis -

PS Claim 15; Fig 1; 768pp; English.

XX The present sequence is human METH1 (ME for metalloprotease and TH for thrombospondin). METH1 can be used for inhibiting angiogenesis in an individual, and for treating cancer, benign tumours, an ocular angiogenic disease, rheumatoid arthritis, psoriasis, delayed wound healing, nonunion endometriosis, vasculogenesis, granulations, hypertrophic scars, nonunion fractures, scleroderma, trachoma, vascular adhesions, myocardial angiogenesis, coronary collateralis, cerebral collateralis, arteriovenous malformation, ischaemic limb angiogenesis, Osler-Webber syndrome, plaque neovascularisation, telangiectasia, haemophilic joints, angiofibroma, fibromuscular dysplasia, wound granulation, Crohn's disease or atherosclerosis. METH1 can also be used in birth control. METH1 can also be used in diagnostic methods for the prognosis of cancer.

XX Sequence 950 AA;

Query Match 52.4%; Score 1960.5; DB 22; Length 950;
Best Local Similarity 53.6%; Pred. No. 4.7e-145;
Matches 388; Conservative 91; Mismatches 178; Indels 67; Gaps 17;

QY 1 MLLGILTLAFAGTAGGFEPEREVVPVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
DB 19 LULLAALLAVSDALGRPSEDEELVVP-ELE-----RAP---GHGTRRLRHAF 64
QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRCFYSGDYNAPDSF 114
DB 65 DQQLDLRLPDSFLAPGFTLQNGRKSGSETPLP-----ETDLAHCFSYGVNGDPSSA 119
QY 115 AAVSLCGGLRGAFYGRGAEYVISPPLNAS---APAAORNSQGA-----HLIQ---RRGVPG 165
DB 120 AALSCEGVRAFYLLGAYFIQPLPAASERLATAAAGEKPPAPLOPQLHLLRRNRQDVG 179
QY 166 PSG-----DPTSRC-----GVASG---WNPAIRLALDPYKPRRAGFSGSRRRS 207
DB 180 TCGVVDDPRPTGKAETDEDEGTEGEGPQWS-----PODPALQGVGP-TGTGS 230
QY 208 GRAKFSVTPRVETFLVADSWKFKCADLEHYLLTLATAARLYRHPILNPIVUV 267
DB 231 IRKKFVSHRYETFMVLVADQSMFAEFGSLGKHYLLTLFVAAKLYKHPISRVSLVV 290
QY 268 KYLLLRDRSDGPKVTGNAAFLRNFCAWQKKNKVSOKHPEYMDTALITRQDLGCATT 327
DB 291 KILVTHDEQKPEVTSNAALTLRNFNCWKQKHNPPSDRDAEHVDTALITRQDLGSGQTC 350
QY 328 DTGLMADVTGMDPKRSCVIEDGLPSAFTTAHELGHVFNMPHDNVKVCVEEYFGKLRAN 387
DB 351 DTGLMADVTGMDPKRSCVIEDGLPSAFTTAHELGHVFNMPHDNVKVCVEEYFGKLRAN 410
QY 388 HMASPTLIQIDRANPWSACSAIITDFLDGSGHGDCLDQPSKPISLPDLPGASYTLISQ 447
DB 411 HMASMLNLDHSPWSPCSAYMITSFLDNGHGECLMDKQNPQLPGLDPTSYDANRQ 470
QY 448 CELAFGVSGKPCP-YMQYCTKLWTKTRAKQOMVQOTRHFHWADGTSGBGKCLKACVCE 506
DB 471 COFTFEGDSKPCDPAASTCTLMWCTGSGGVIVQCTHFHWADGTSGBGKCLKACVCE 530
QY 507 RNLANKH---RVDSGWAKWDYGPCSRCTCGGVGVOLARQCTNPTTPANGKVCGRVVKYR 563
DB 531 KTD-RKHEDTFPHGSGWGWGFWGDCSRCTCGGVGVYTHRECDNVPVKNNGKYCEGRVRYR 589
QY 564 SCNLEPCPSSASGKSFREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDKCKLICRAN 623
DB 590 SCNLEDCPDN-NGKTFREEQCEAHNEFSKASFGSGPAVEMTPKYAGVSPKDRCKLICQAK 648
QY 624 GTGYEYVLAPKVDGTLCSPOSTSVCGQKCIKAGCDNGLSKRRFDKCVCGGDNKSK 683
DB 649 GIGYEFVLQPKVVDGTLCSPOSTSVCGQKCIKAGCDNGLSKRRFDKCVCGGDNKSK 708
QY 684 KVTG 687
|::|

DB 709 KISG 712

RESULT 10

AAW80285
ID AAW80285 standard; Protein; 967 AA.
XX AC AAW80285;
XX 19-JAN-1999 (first entry)
XX Human integrin ligand polypeptide ITGL-TSP.
XX ITGL-TSP; integrin ligand; angiogenic disease; cancer; atherosclerosis;
XX chronic inflammatory disorder; rheumatoid arthritis; tissue remodeling;
XX macular degeneration; diabetic retinopathy; Alzheimer's disease; human;
XX restenosis.
XX OS Homo sapiens.
XX PN EP874050-A2.
XX 28-OCT-1998.
XX 27-JAN-1998; 98EP-0300575.
XX 24-APR-1997; 97US-0845496.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA (SMIK) SMITHKLINE BEECHAM CORP.
XX PA (SMIK) SMITHKLINE BEECHAM PLC.
XX Fronwald JA, Hastings GA, Jonak ZL, Terrett JA;
PI Trull SH;
XX WPI; 1998-544643/47.
XX N-PSDB; AAV66508.
XX DNA encoding integrin ligand polypeptide ITGL-TSP - used to treat
XX angiogenic diseases, restenosis, Alzheimer's disease and in tissue
XX remodeling
XX Claim 11; Pages 6-9; 24pp; English.
XX This represents a human integrin ligand polypeptide ITGL-TSP. ITGL-TSP
XX polypeptides can be used in the treatment of angiogenic diseases such as
XX cancer, cancer metastasis, chronic inflammatory disorders, rheumatoid
XX arthritis, atherosclerosis, macular degeneration or diabetic retinopathy,
XX restenosis, Alzheimer's disease and tissue remodeling. They can be used
XX to treat a subject in need of enhanced activity or expression of the
XX ITGL-TSP polypeptide.
XX Sequence 967 AA;
Query Match 52.4%; Score 1960.5; DB 19; Length 967;
Best Local Similarity 53.6%; Pred. No. 4.8e-145;
Matches 388; Conservative 91; Mismatches 178; Indels 67; Gaps 17;
QY 1 MLLGILTLAFAGTAGGFEPEREVVPVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
DB 36 LULLAALLAVSDALGRPSEDEELVVP-ELE-----RAP---GHGTRRLRHAF 81
QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRCFYSGDYNAPDSF 114
DB 82 DQQLDLRLPDSFLAPGFTLQNGRKSGSETPLP-----ETDLAHCFSYGVNGDPSSA 136
QY 115 AAVSLCGGLRGAFYGRGAEYVISPPLNAS---APAAORNSQGA-----HLIQ---RRGVPG 165
DB 137 AALSCEGVRAFYLLGAYFIQPLPAASERLATAAAGEKPPAPLOPQLHLLRRNRQDVG 196
QY 166 PSG-----DPTSRC-----GVASG---WNPAIRLALDPYKPRRAGFSGSRRRS 207
DB 197 TCGVVDDPRPTGKAETDEDEGTEGEGPQWS-----PODPALQGVGP-TGTGS 247

QY 208 GRAKRFVSIPIRYVETLVVADESMVKHGADEHYLLTLATAARLYRHPSILNPIVIVV 267
Db 248 IRKRFVSSHRYVETMLVADQSAEFGSLKHYLLTFSVAARLYRHPSIRNSVLVVV 307
QY 268 KVLRLDRDSDGPKVTGNAALTLRNFCAWKKLNKSKHPEYWDTAILFTRODLGCGATTC 327
Db 308 KILVIHDEQKGPVTSNAALTLRNFCAWKKLNKSKHPEYWDTAILFTRODLGCGATTC 367
QY 328 DTLMGADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEFGKLRAN 387
Db 368 DTLMGADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEFGKLRAN 427
QY 388 HMASPTLIQIDRANPWSACSAIITDFLDGSHGDCILLDOPSKPISLPEDLPGASYTLSSQ 447
Db 428 HMASPTLIQIDRANPWSACSAIITDFLDGSHGDCILLDOPSKPISLPEDLPGASYTLSSQ 487
QY 448 CELAFGVGSKPCP-YMOYCTKLWCTGKAKQMVQCTRHFPWADGTCGEGSKLCKAGACVE 506
Db 488 CQFTFGEDSKHCPDAASTCTLWCTGTSGVLVCQTKHFPWADGTCGEGSKLCKAGACVE 547
QY 507 RHNLNKH---RVDSWAKWDPYPCSTCGGVLQARRQCTNTPANGGKYCBGVRYKVR 563
Db 548 KTD-RKHFDPFPGSGWGMGFWDCSTCGGVLQARRQCTNTPANGGKYCBGVRYKVR 606
QY 564 SCNLEPCPSASGSKFRECEAFNGYNHSTNLTILAVAWPKYSGVSPRDKCKLICRAN 623
Db 607 SCNLEPCPDN-NGKTFRECEAHNFSKASFGSGFAVEWPKYSGVSPRDKCKLICRAN 665
QY 624 GTGYFVYVLPKVVVDGTLCPDSTSVQVQKICIRAGCDGMLGSKRRFDKCGVCGDNKSK 683
Db 666 GIGYFVYVLPKVVVDGTLCPDSTSVQVQKICIRAGCDGMLGSKRRFDKCGVCGDNKSK 725
QY 684 KVTG 687
Db 726 KISG 729

RESULT 11
AAB50011
ID AAB50011 standard; Protein; 968 AA.
XX AC AAB50011;
XX DT 19-MAR-2001 (first entry)
XX DE Protein; SEQ ID 125.
XX KW Human; METH1; metalloprotease; thrombospondin; angiogenesis inhibition;
KW cancer therapy; benign tumour; ocular angiogenic disease;
KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;
KW coronary collateral; cerebral collateral; arteriovenous malformation;
KW ischaemic limb angiogenesis; Osler-Webber syndrome;
KW plaque neovascularisation; telangiectasia; haemophilic joint;
KW angiofibroma; fibromuscular dysplasia; wound granulation;
KW Crohn's disease; atherosclerosis; birth control.
XX OS Homo sapiens.
XX PN WO200071577-A1.
XX PD 30-NOV-2000.
XX PF 25-MAY-2000; 2000WO-US14462.
XX PR 25-MAY-1999; 99US-0318208.
PR 20-JUL-1999; 99US-0144882.
PR 10-AUG-1999; 99US-0147823.
PR 13-AUG-1999; 99US-0373658.
PR 22-DEC-1999; 99US-0171503.
PR 22-FEB-2000; 2000US-0183792.

XX PA (HUMA-) HUMAN GENOME SCI INC.
PA (SMK) SMITHKLINE BEECHAM CORP.
PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
PA (IRUE/) IRUELA-ARISPE L.
PA (HAST/) HASTINGS G A.
PA (RUB/) RUBEN S M.
PA (JON/) JONAK Z L.
PA (TRUL/) TRULLI S H.
PA (FORN/) FORNWALD J A.
PA (TERR/) TERRETT J A.
PI IrueLA-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
PI Fornwald JA, Terrett JA;
XX WPI: 2001-025136/03.
XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
PT rheumatoid arthritis and psoriasis -
XX Claim 15; Pages 759-763; 768pp; English.
XX The present invention relates to human METH1 and METH2 (ME for
CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
CC METH can be used for inhibiting angiogenesis in an individual, and for
CC treating cancer, benign tumours, an ocular angiogenic disease,
CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
CC fibromuscular dysplasia, wound granulation, Crohn's disease or
CC atherosclerosis. METH can also be used in birth control. METH can also
CC be used in diagnostic methods for the prognosis of cancer. The present
XX sequence is a protein isolated in the present invention.
XX Sequence 968 AA;
Query Match 52.4%; Score 1960.5; DB 22; Length 968;
Best Local Similarity 53.6%; Pred. No. 4.8e-145;
Matches 388; Conservative 91; Mismatches 178; Indels 67; Gaps 17;
QY 1 MLLGILLTAPAGTAGGFEPEVVPVIRLDPDINGRYYWRGPEDSGDGLFQITAF 60
Db 37 LLLAAALLAVSDALGRPSEDEELVVP-ELE-----RAP---GHGTTTLRLHAF 82
QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRCFYSGDVNAEPDSF 114
Db 83 DQQLDLRLPDSSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCIFYSGTVNGDPSSA 137
QY 115 AAVSLCGGLGAGFGVRYGAEYVISPLPNAS---APAAQNSOGA-----HLLQ--RGVPGG 165
Db 138 AALSCEGVGAFYLLGEAYFIQPLPAASERLATAAPGKPPAPLQFHLRLNRQGVGG 197
QY 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPVKPRAGGSESRSSRS 207
Db 198 TCGVVDDEPRPTGKAETEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGEDEGE 248
QY 208 GRAKRFVSIPIRYVETLVVADESMVKHGADEHYLLTLATAARLYRHPSILNPIVIVV 267
Db 249 IRKRFVSSHRYVETMLVADQSAEFGSLKHYLLTFSVAARLYRHPSIRNSVLVVV 308
QY 268 KVLRLDRDSDGPKVTGNAALTLRNFCAWKKLNKSKHPEYWDTAILFTRODLGCGATTC 327
Db 309 KILVIHDEQKGPVTSNAALTLRNFCAWKKLNKSKHPEYWDTAILFTRODLGCGATTC 368
QY 328 DTLMGADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEFGKLRAN 387
Db 369 DTLMGADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEFGKLRAN 428
QY 388 HMASPTLIQIDRANPWSACSAIITDFLDGSHGDCILLDOPSKPISLPEDLPGASYTLSSQ 447

Db	429	HMWASMLNSLDHSPSPCSAYMTISFLDNGHGECIMDKPQNPIQLPGDLPETSYDANRQ	488
Qy	448	CELAFGVGSKPCP-YMQYCTKLMCTGKAKGQMVQCOTRHFPWADGTSCEGKLCCLKGACVE	506
Db	489	CQTFEGEDSKHCPDAASTCSTLWCTGTSGVLVCQTKHFPWADGTSCEGKWCINGKCVN	548
Qy	507	RHNLNKH--RVDGSAKWADPYGCSRTCGGGVQIARQCNTNPTPANGGKYCEGVRYKYR	563
Db	549	KTD-RKHEDTTPFHGSGMGPWGDCSRTCGGGVQYTMRECDNPVKNGGKYCEGKRVYR	607
Qy	564	SCNLEPCPSASGKSPREEQCEAFNGYNHSTNRLTLAVAWVPKYSVSPDRCKCLICRAN	623
Db	608	SCNLEDCPDN-NGKTFREEQCEAHNEFSKASFGSGPAVEWIPKYAGVSPDKCKLICQAK	666
Qy	624	GTGYFVYLPAPKVVDTGLCSFDSSTVCVQGGKCIKACDGNLGSKKRFDKCGVCGGDNKSCK	683
Db	667	GIGYFVFLQPKVVDGTPCSPDSTSVCVQGCYKAGCDRIIDSKKKFDKCGVCGGNGSTCK	726
Qy	684	KVTS 687	
Db	727	KISG 730	
RESULT 12			
AAY04142			
ID	AAY04142 standard; Protein; 967 AA.		
AC	AAY04142;		
XX			
DT	15-JUN-1999 (first entry)		
XX			
DE	Human Tango-71 protein.		
XX			
XX	Human; Tango-71; Tango-73; Tango-74; Tango-76; Tango-83; diagnosis;		
KW	detection.		
XX	Homo sapiens.		
OS	W09907850-A1.		
PN	18-FEB-1999.		
XX			
XX	06-AUG-1998; 98WO-US16502.		
PF			
XX			
PR	05-SEP-1997; 97US-0058108.		
PR	06-AUG-1997; 97US-0054966.		
XX			
XX	(MILL-) MILLENNIUM BIOTHERAPEUTICS INC.		
PA			
XX			
PI	Goodearl ADJ, Holtzman DA;		
XX			
PI	WPI; 1999-167426/14.		
DR	N-PSDB; AAX19955.		
XX			
PT	New TANGO polypeptides and nucleic acids encoding them - useful as		
PT	diagnostic agents and for treating disorders caused by aberrant		
PT	expression of TANGO		
XX			
PS	Claim 8; Fig 1; 84pp; English.		
XX			
CC	The present sequence represents human Tango-71. Tango polypeptides are		
CC	useful for identifying compounds which bind the polypeptide via direct		
CC	binding, competition binding assays or Tango-71, -73, -74, 76 or -83-		
CC	mediated signal transduction. Tango polypeptides are also useful for		
CC	identifying modulating compounds by determining effect on Tango activity.		
CC	Tango polypeptides and nucleic acids are useful for diagnosing diseases		
CC	related to aberrant expression of Tango, and Tango polypeptides are		
CC	useful for raising antibodies which can be used in diagnostic assays for		
CC	detection of Tango, and also for generating anti-idiotypic antibodies for		
CC	prevention and protection.		
XX	Sequence 967 AA;		
SQ			

CC The present sequence represents a murine ADAMTS-1 protein. ADAMTS-1 is
CC a metalloproteinase. The specification describes another related
CC metalloproteinase, a C. elegans GON-1 protein, that lacks a transmembrane
CC domain and possesses a predicted metalloprotease domain between residues
CC 269-456. In C. elegans hermaphrodites, GON-1 is required for migration of
CC two distal tip cells to produce elongated tubes, whereas in males, GON-1
CC is required for migration of a single linker cell to produce a single
CC elongated tube. The protein is used in the method of the invention. The
CC specification describes a method for identifying a modulator of a
CC protein that contains a metalloproteinase domain and a thrombospondin
CC domain. The method comprises treating a target organism, having a test
CC developing gonadal cell that is responsive to the protein, with a test
CC compound, and determining any change in migration or shape of the cell
CC attributable to the test compound. The compounds identified are
CC potential therapeutic modulators of abnormal cell migration and organ
CC shaping, e.g. for rendering animals (specifically nematodes) sterile
CC and for inhibiting cancer metastases.

XX
SQ Sequence 950 AA;

Query Match 51.3%; Score 1921; DB 21; Length 950;
Best Local Similarity 52.0%; Pred. No. 5.9e-142;
Matches 379; Conservative 95; Mismatches 177; Indels 78; Gaps 17;

QY 1 MLLGLTLTAFAGTAGG--PEPEEVVPIRLDPDINGRYYWRGP-EDSGDQGLIFQI 57
DB 20 LULLASITMLLCARGAGRTEDEELVLP-SLE-----RPGHDSITTRL--RL 66
QY 58 TAFQEDFVHLTPDQAFLAPAFSTHGLVPLQGLTGGS-----SDLRRCFYSGDVNA 109
DB 67 DAFGQQLHLKLPDSDSFLAPGFTLTQV----GRSPGEAQLDPTGDLAHCIFYSGTVNG 121
QY 110 EPDSPAANVSLCGLRGARGYGAERVISPLPNAS-----APAAQRNSOGA----HLLORRG 161
DB 122 DFGSAALSLCGYRGATYLOGEFFIOPAFGVATERLAPAVPEESARPPQHILRRR- 180
QY 162 VPGPGSDPTSCRGVAGSNPAILRALDPYKPRAGFESRRR----- 206
DB 181 -----RRSGGKCGYMD-----DETLPDSRPESONTNQPVRDPTPDAGKP 226
QY 207 -----SGRAKRFVSPRYVETLVVADESVMKFGADLEHYLLTLTAARLYRHPISLNP 262
DB 227 SGPGSIRKKRFVSPRYVETMLVADQSMADFHSGSLHYLLTLFVSAARFYKHPISIRNS 286
QY 263 NIWVKVLLLRDSDGKPVGTGNAALTNRNCAWOKLNKVSDDKPEYWDTAILETRQDLC 322
DB 287 SLVVKVILVIEEQGPEVTSNAALTNRNCSWQKOHNSPSDRDPEHYDTAILETRQDLC 346
QY 323 GATCTDLGMADVTGMDPKRCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCCEVFG 382
DB 347 GSHTCDTLGMADVTGMDPSRCSVIEDDG-QAAFTTAHELGHVFNMPHDNKAHCASLNG 405
QY 383 KLRANHMSPPTLIQIDRANPWSACSAAITDPLDSHGDCLLDOPSKPISLPEDLPGASY 442
DB 406 VSGDHLMSMLSSLDHSPWPCSAIYVTFDLNMGHGECLMDKPNPIKLPDLPGLTLY 465
QY 443 TLSQQCELAFGVSKPCP--YMOYCTKLMCTGKAKGMVQCTRHFPWADGTSCGEGKLC 501
DB 466 DANROCCFTFGEESKHCPCDAATCTTLCCTGTSGLLVCCCTKHPWADGTSCGEGKWCVS 525
QY 502 GACVERHNLNKH---RVDSWAKWDYPGPCSRTCGGGVQLARRQCTNPTPANGKCYCEGV 558
DB 526 GKCVNTDM-KHFATPVHGSWGPWGPWGDSCSRTCGGGVQYTMRECDNPVPKNGKCYCEGK 584
QY 559 RVKYRSCNLEPCPSSASGKSFREEQCEAFNGYNHSTNRLTLAVAWVPKYSGVSPDRCKL 618
DB 585 RVKYRSCNIEDCPDN-NGKTFREEQCEAHNEESKASFCGNEPTVEWTPKYAGVSPDRCKL 643
QY 619 ICRANGTGYVILAPKVVDTGLCSDDSTSVQVQKICVAGCDGNLGSKKRDKCGVCGGD 678
DB 644 TCEAKGIGYFFVLPKVVDTGPCSPDSTSVQVQGVQKAGCDRIIDSKKFKDKGCVCGGN 703
QY 679 NKSCCKKVTG 687

Db 704 GSTCKKMSG 712
Search completed: April 29, 2003, 17:20:36
Job time : 36.4317 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 11.1439 Seconds
(without alignments)
1813.869 Million cell updates/sec

Title: US-10-009-332-1_COPY_1_687

Perfect score: 3744
Sequence: 1 MLLGILTLAFAGTAGGFE.....RPDKGCGGDNKSKKVTG 687

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1960.5	52.4	967	4	US-09-130-491-2
2	1689.5	45.1	837	4	US-09-122-126B-2
3	1679	44.8	905	4	US-09-369-364A-9
4	1611	43.0	930	4	US-09-369-364A-2
5	1587.5	42.4	930	4	US-09-122-126B-15
6	1536	41.0	608	4	US-09-130-491-13
7	1432	38.2	1882	4	US-09-369-364A-13
8	1394.5	37.2	874	4	US-09-369-364A-15
9	1077	28.8	997	4	US-09-369-364A-7
10	1072	28.6	551	4	US-09-130-491-16
11	1004	26.8	518	4	US-09-369-364A-22
12	976	26.1	1211	4	US-09-491-522-5
13	968.5	25.9	1224	4	US-09-930-872-4
14	944.5	25.2	1205	4	US-09-491-522-11
15	930.5	24.9	1081	4	US-09-369-364A-17
16	905.5	24.2	859	4	US-09-369-364A-5
17	841	22.5	481	4	US-09-130-491-8
18	662.5	17.7	245	4	US-09-369-364A-11
19	576.5	15.4	566	4	US-09-491-522-7
20	441	11.8	491	4	US-09-930-872-2
21	358	9.6	525	4	US-09-369-364A-21
22	346.5	9.3	812	4	US-09-632-098-4
23	339	9.1	802	4	US-09-632-098-2
24	303	8.1	464	4	US-09-411-329C-14
25	301.5	8.1	616	4	US-09-608-790-1
26	301	8.0	462	4	US-09-411-329C-3
27	301	8.0	462	4	US-09-411-329C-17

28	301	8.0	529	2	US-08-836-442-3	Sequence 3, Appli
29	285	7.6	621	4	US-09-026-001A-6	Sequence 6, Appli
30	278	7.4	621	4	US-09-026-001A-18	Sequence 18, Appli
31	277.5	7.4	814	4	US-09-813-819-4	Sequence 4, Appli
32	277.5	7.4	814	4	US-09-920-048-4	Sequence 2, Appli
33	277.5	7.4	855	4	US-09-813-819-2	Sequence 4, Appli
34	277.5	7.4	855	4	US-09-920-048-2	Sequence 2, Appli
35	273	7.3	613	4	US-09-026-001A-10	Sequence 10, Appli
36	269.5	7.2	769	1	US-08-243-542-4	Sequence 4, Appli
37	269.5	7.2	769	1	US-08-477-407-4	Sequence 4, Appli
38	269.5	7.2	769	1	US-08-484-355-4	Sequence 4, Appli
39	259	6.9	592	4	US-09-026-001A-14	Sequence 14, Appli
40	253.5	6.8	670	1	US-08-243-542-3	Sequence 3, Appli
41	253.5	6.8	670	1	US-08-477-407-3	Sequence 3, Appli
42	253.5	6.8	670	1	US-08-484-355-3	Sequence 3, Appli
43	248.5	6.6	1170	1	US-08-313-288B-20	Sequence 20, Appli
44	244	6.5	335	4	US-09-152-060-64	Sequence 64, Appli
45	241.5	6.5	751	2	US-08-836-443-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-130-491-2
; Sequence 2, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130.491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-130-491-2

Query Match	52.4%	Score 1960.5;	DB 4;	Length 967;
Best Local Similarity	53.6%	Pred. No. 9.6e-156;		
Matches	388;	Conservative	91;	Mismatches 178;
				Indels 67;
				Gaps 17;
QY	1	MLLGILTLAFAGTAGGFEPEPEVVPVIRLDPDINGRRYWRGPDSDGGLFQITAF	60	
DB	36	LULLAALLAVSDALGRFSEDEELVVP-ELE-----RAP---GHGTTLRRLHAF	81	
QY	61	QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRRCFYSGDVNAEPDSF	114	
DB	82	DQOLDLELRPDSFAPGFTLQNVGRKSGSETPLP-----ETDLAHCYFSTVNGDPSA	136	
QY	115	AAVSLCGGLRGAFGRGAEYVISPFPNAS---APAAQNSOGA---HLLQ--RRGVPG	165	
DB	137	AALSCEGVGRGAFYLLGEAYFTQPLPAASERLATAAPGKPPAPLQFHLRRNQDVG	196	
QY	166	PSQ-----DPTSRC-----GVASG--WNPAILRALDPPKPRAGFSGSRRS	207	
DB	197	TCGVVDDEPRPTKGAETEDGTEGEGPOWS-----PQDPALQGVGP-TGTGS	247	
QY	208	GRAKREVSIPRYVETLWVADESMVKFHGADLEHYLLTLATAARLYRHPSTLNPIVIV	267	
DB	248	IRKREVSISHRYVETMLVADQSMAEFHGSLKHYLLTLFVSAARLYKHPSTRNSVSLVV	307	
QY	268	KVLLLRDRSGPKVTCNAALTFRNFCANWKKLNKYSDKHPEYWDPAIFLTRQDLGGATTC	327	
DB	308	KILVTHDEQGPVTSNAALTFRNFCANWKKLNKYSDKHPEYWDPAIFLTRQDLGGATTC	367	

QY 328 DTLMADVGTMCOPKRCSCSVIEDDGLPSAFTTAHELGHVFNPHNDNVKVCVEFGKLRAN 387
Db 368 DTLMADVGTMCOPKRCSCSVIEDDGLQAAFTTAHELGHVFNPHNDNVKVCVEFGKLRAN 427
QY 388 HMMSPITLIQIDRANPWSACSAIIIDFDSHGDCLLDQPSKPSLPELPGASYTISQO 447
Db 428 HMMASMLNLDHSQWSPSCSAIYITSLFNDNGHGECLMDKPNQITLPGDLPSTYDANRQ 487
QY 448 CELAFGVGSKPCP-YMYCTKLTWCTGKAGOMVOCOTRHPWADGTSCGEGKCLKLGACVE 506
Db 488 CQTFEGEDSKHCPDAASTCTWCTGTSGVLVCQTKHPWADGTSCGEGKWCINGKCNV 547
QY 507 RHLNKH---RVDGSMWAKWDPGSCSRTCGGVQVQLARQCTNPTPANGKCYCEGVRYKYR 563
Db 548 KTD-RKHEDTPEHSGKMGWPGWDCSRTCGGVQVYTMEDCNVPYKNGKCYCEGVRYR 606
QY 564 SCHLEPCPSAGSKSFREOQCAFNGYNHSTNRLTLAVAWPKYSGVSPDRCKLICRAN 623
Db 607 SCHLEDCPDN-NGKTFREOQCAHNEFSKASFGSGPAVEWIPKVGYSKDRCKLICQAK 665
QY 624 GTGYFYVLAPKVVVDGTLCSPDSTSCVQVKCIKAGCDNGLSKKRFDKCGVCGDGNKSK 683
Db 666 GIGYFVVLQPKVVVDGTLCSPDSTSCVQVKCIKAGCDNGLSKKRFDKCGVCGDGNKSK 725
QY 684 KVTG 687
Db 726 KISG 729

RESULT 2

US-09-122-1268-2

; Sequence 2, Application US/091221268

; Patent No. 6451575

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: AGGREGAN DEGRADING METALLO PROTEASES

; FILE REFERENCE: DM6909

; CURRENT APPLICATION NUMBER: US/09/122.1268

; CURRENT FILING DATE: 1998-07-24

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 2

; LENGTH: 837

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-122-1268-2

Query Match 45.1%; Score 1689.5; DB 4; Length 837;

Best Local Similarity 47.2%; Pred. No. 4e-133;

Matches 328; Conservative 103; Mismatches 215; Indels 49; Gaps 14;

QY 1 MLLLGTLTAFAGRTAGGFEPEREVVPVIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
Db 37 LLLLLLTLASLPARLASLPREEIIVFEKLVSVL-----PGSGAPARLLCRLQAF 88
QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVLPQGLTGSSDLRCFCYSGDVNAEPDSFAAYSLC 120
Db 89 GETLLELRQDSGVQVQVGLTVQVLAQPE-LLGGAEP--GTYLGCTINGDPESVASLHWD 145
QY 121 GG-LRGAFCYRGAEEVVISPLPNASAPAAQRNSQGAHLRQGVGGSDPTSRCGVASG 179
Db 146 GGALLGVQYRGAELHQLPQEGGTNSA--GGPGAHLLRRK-----SPASGGPNCN--- 196
QY 180 WNPAILRALDPKPRAGGSRSSRRSRRAKRFVSPRYVETILVWADSVKHFAGDL 239
Db 197 -----KAPLSPSPRPR--RAKRFASLSRFVETILVWADSKAAHFGAGLK 239
QY 240 HYLLLTATAARLYRHPISLIPINIVVVKVLLLRDRDSGPKVTGNAALTIRNFCAWOKKL 299
Db 240 RYLLTVMMAAAKAFKHPISIRNPVSLVTRVLILGSGEGPQVGPSPAAQTLRSFCAWORGL 299
QY 300 NKVSKHPEYWDTAILETRQDLGATTCTDTLGMADVGTMCDPKRCSCSVIEDDGLPSAFTT 359

Db 300 NTPESDPPHFTAILFTRQDLGCVSTCDTLMADVGTVCDPARSCAIVDDGQLQSFTA 359
QY 360 AHELGHVFNPHNDNVKVCVEFGKL-RANHMSPTLIQIDRANPWSACSAIITDFLDSG 418
Db 360 AHELGHVFNPHNDNVKVCVEFGKL-RANHMSPTLIQIDRANPWSACSAIITDFLDSG 418
QY 419 HGCDLLDQPSKPSLPELPGASYTISQOCELAFGVGSKPCPYM-QYCTKLWCTGKAKQ 477
Db 420 YGCHCLDRPEALHLPVTFPGKDYDADQCOLTFGPSRHCPLPPPCALWCSGHLNGH 479
QY 478 MYCQTRHFPWADGTSCGEGKCLKAGACVERHNLNKHRYD--GSWAKWDPYGPCSRTCGG 535
Db 480 AMCQTKHSPWADGTSCGEGKCLKAGACVERHNLNKHRYD--GSWAKWDPYGPCSRTCGG 539
QY 536 VOLARROCTNPTPANGKCYCEGVRYKYSCLNLEPCPSSASCKSFREOQCAFNGYNHSTN 595
Db 540 VOFSSRDCTRPVPRNGGKYCEGRTRFRSCNTEPCP-TGSALTTFREOQCA---YNHRTD 595
QY 596 RLTL---AVAWPKYSGVSPDRCKLICRANGTYFYVLAPKVVVDGTLCSPDSTSCVQ 652
Db 596 LFKSPGMDWVPRTYGVAPDOCKLTQARALGYVYVLEPRVVDGTPCSPDSSSVQVQ 655
QY 653 KCIKAGCDNGLSKKRFDKCGVCGDGNKSKKVTG 687
Db 656 RCIHAGCDRIIGSKKFKDKMVCVGGDGGSCSKQSG 690

RESULT 3

US-09-369-364A-9

; Sequence 9, Application US/09369364A

; Patent No. 6391610

; GENERAL INFORMATION:

; APPLICANT: Apte, Suneel

; APPLICANT: Hurskainen, Tiina L.

; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases

; FILE REFERENCE: 26473/4007/10-30-00

; CURRENT APPLICATION NUMBER: US/09/369.364A

; CURRENT FILING DATE: 1999-08-06

; NUMBER OF SEQ ID NOS: 31

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 9

; LENGTH: 905

; TYPE: PRT

; ORGANISM: Mus musculus ADAMTS-8

US-09-369-364A-9

Query Match 44.8%; Score 1679; DB 4; Length 905;

Best Local Similarity 46.7%; Pred. No. 3.4e-132;

Matches 344; Conservative 101; Mismatches 202; Indels 90; Gaps 18;

QY 2 LLLGLTLT-----APAGRTAGGFEPEREVVPVIRLDPDINGRRYWRGPDSDGQGL 53
Db 13 LLLLLLQPPPLVCGAPAGPTGA--QASELVVPTRL-----PGSASE--L 55
QY 54 IFQITAFQEDFYHLTPDAQFLAPAFSTEHLGVLPQGLTGSSDLRCFCYSGDVNAEPDS 113
Db 56 AFHLSAFQGGFVLRAPDASFLAPEFKIERLG-GSSAAGGEPGLRCGFSGVNGERES 114
QY 114 FAANSLCGLRCGAFGYRGAEEVVISPLPNASAPAAQRNSQGA-----HLLQRGVPGGP 166
Db 115 LAAMSCVAGWSGSLFLLAGEEFTIQP-----QGAGDSLDPQPHLRQWG-PGOR 160
QY 167 SCDPTSRCGVASGWNPAITRALDPKPRAGGSRSS-----RRRSR----- 208
Db 161 REDP---GLAAAEVFPPLQGLEWEVEVNGQOGRSDNEEDKKQDEGLLKETEORKV 216
QY 209 -----RAKRFVSPRYVETILVWADSVKHFAGDLHYLLTATAARLYRHPISILN 260
Db 217 PPPFGSKTRSKRFVSEARFETLLVADASMAAFYGTDLQNHILTVMSMAARIYKHPISIRN 276
QY 261 PINIVVVKVLLLRDRDSGPKVTGNAALTIRNFCAWOKKLNVSKHPEYWDTAILETRQD 320

```

Db 277 SYNLVVKVLIYKERWGPVSDNGGLTLRNFCSKORFNKPSRHEHYDTALLTRQN 336
QY 321 LCG-ATTCDTLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKKVCEE 379
Db 337 FCGKEQCQDTLGMADVTGICDPDKSCSVIKDEGLQAAAYTLAHELGHVLSMPHDDSKPCVR 396
QY 380 VFGKLRANHMSPPTLIQIDRANPWSACSAIIITDFLDSHGDCDCLLDOPSPISLPEDLPG 439
Db 397 LFGPMCKYHMAFFIHNKTLNPSPCSAVYLTLLDDHGDCDCLLDAPTSVLPPTGLPG 456
QY 440 AS--YTLSSQCEBLAFVSGSKPCP---YMOYCTKLWCTGKAKQMWVCOTRH--FPWADGTS 492
Db 457 HSTLYELDOQCQIFGPDPRHCNPTSVEDICVOLCARHDSDEPCHTKNGSLILWADGP 516
QY 493 CGEGKLCIKAGACVERHNLN--KHRVDSWAKWDVPGCSRTCCGGVGLAROCNTNTPAN 550
Db 517 CGPGHCLDGSCLVKEDVENKAVVDGWPWRWGQSRCTCCGGIGTFSNRECDNMPQN 576
QY 551 GKGYCEGVRYKRYSCNLEPCPSSASGSKSFREEOCEAFNGYNHSTNRLTLAVAWVPKYSV 610
Db 577 GGRFCLGERVKYQSCNTEBEP--PNGKSFREOQCEKYNAVNH--TDLGDNFLQWVPKYSV 633
QY 611 SPRDCKLICRANGTYFVLPAPKVVVDGTLCSPDSTSVCOGKCIKAGCDNLSKKRFD 670
Db 634 SPRDCKLCFRARGSEKVFPAKVIDGTLCPDTLSCVVRGQCVKAGCDHVNSPRKL 693
QY 671 KCGVCGGDNKSKCKVTG 687
Db 694 KCGVCGGKTACKRISG 710

```

RESULT 4

```

US-09-369-364A-2
; Sequence 2, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 930
; TYPE: PRT
; ORGANISM: mus musculus ADAMTS-5
US-09-369-364A-2

```

```

Query Match 43.0%; Score 1611; DB 4; Length 930;
Best Local Similarity 47.3%; Pred. No. 1,7e-126;
Matches 305; Conservative 82; Mismatches 182; Indels 76; Gaps 7;

QY 90 GLTGGSSDLRRFYSGDVNAEPDSPAIVSLCGGLRGAFYGAAYVISPLNAPSAPAAQR 149
Db 121 GLSASSGHRGHCYRGTVDGSPRSLAVFDLCGLDGFFAVKHARYTLKPLLRGWSAEYER 180
QY 150 ----NSOGAHLQRRG-----VPGPSPGDPTRSGCVASGWNPAIRLALDPY 191
Db 181 IYDGGSSRIILVYNRGFSFALPPRASCETPASFSG----- 217
QY 192 KPRAGFGESRRRRSG-----RAKRFVSIPIRYVETLVVAD 227
Db 218 -POESPVSRSRRRSALAPQLLDHSAFSPGNAGPQTWRRRRRSISRARQVELLLVAD 276
QY 228 ESMVKFHADLEHYLTLLATARYRHSILNPNIVVVKVLLLRDSDGPKYTGNAAL 287
Db 277 SSARMYGLQHYLLTMASIANRYSIASIENHRLAVKVVVVDTKDTSLEYSKNAAT 336
QY 288 TLRFCAWOKLKNVSKDHPYWDTAILTRDLCGATTCDTLGMADVTGMDPKRSCSV 347

```

```

Db 337 TLKNECKWQHQNOLGDDHDEHYDAAILFTREDLCGHHSCTDLGMADVGTICSPERSCAV 396
QY 348 IEDDGLPSAFTTAHELGHVFNPHDNVKKVCEEVFGKLRANHMSPPTLIQIDRANPWSACS 407
Db 397 IEDDGLHAAFTVAHEIGHLLGLSHDSDSKCEBENFGTTEDKRLMSSILTSIDASKPWSKCT 456
QY 408 AAIITDFDSDHGDCDCLLDQPSKIPISLPEDLPASVYTLSSQCEBLAFVSGSKPCPYMOTYCTK 467
Db 457 SATITEFDLDDHGNCCLLDLPKQILGPEELPGQYDATQCCNLTFGPPEYSVCPGMDVCAR 516
QY 468 LMCTKAKQMWVCQTRHPFPWADTSCGEGKLCIKAGACVER-----HNLNKHVRVDGSAKW 522
Db 517 LMCAYVRQGMVCLTKLPFAVEGTGCGKGRVCLQKCDKTKKKYKYSTSH--GNGMSW 573
QY 523 DPYGPCSRTCCGGVGLAROCNTNTPANGKYCEGVRYKRYSCNLEPCPSSASGSKSPREE 582
Db 574 GPWGQCSRSCGGGVGFAYRHCNNPAPRNSGRYCTGKRAIYRSCSVTPCP--PNGKSFRE 631
QY 583 OCEAFNGYNHSTNRLTLAVAWVPKYSVSPRDKCKLICRANGTYFVYVLPAPKVVVDGTLCS 642
Db 632 OCEAKNGYQSDAKGVKTFVWVPKYAGVLPADVCKLTCAKGTGYVYVFPKVDGTGTECR 691
QY 643 PDSTSVCOGKCIKAGCDNLSKKRFDKCGVCGGDNKSKCKVTG 687
Db 692 PYSNSVYVRGRCVTRGCDGIGSKLQYDKGCVCGGDNSSCTKIIG 736

```

RESULT 5

```

US-09-122-126B-15
; Sequence 15, Application US/09122126B
; Patent No. 6451575
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: AGGRECAN DEGRADING METALLO PROTEASES
; FILE REFERENCE: DMG909
; CURRENT APPLICATION NUMBER: US/09/122,126B
; CURRENT FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-122-126B-15

```

```

Query Match 42.4%; Score 1587.5; DB 4; Length 930;
Best Local Similarity 47.2%; Pred. No. 1.6e-124;
Matches 322; Conservative 82; Mismatches 209; Indels 69; Gaps 15;

QY 48 SGDQGLIFQITAFQEDFYHLHTPDQFLAPAFSTEHLGVPLQGLTGSSDLR---RCFYS 104
Db 82 SGGGKGVLYVAGRRRFLDLDERDGSVGTAGF-----VPAGG--GTSAPWRHSHCFYR 133
QY 105 GDVNAEPDSPAIVSLCGGLRGAFYGAAYVISPL---PNASAPAAORNSOGA---HLL 157
Db 134 GTVDASPRSLAVFDLCGLDGFFAVKHARYTLKPLLRGFAEENKGRVYGGSGARILHY 193
QY 158 QRRG-----VPGGPS-----GDPTRSGCVASGWNPAIRL--RALDPYKP 193
Db 194 TREGFSFALPPRASCETPASTPEAHEHAPAHNSNPGRALAS---QLLDQSALSP--- 246
QY 194 RRAGFES---RSRRSRAKRFVSIPIRYVETLVVADESVMKFFHCADLEHYLLTLATAA 250
Db 247 -AGGSGPQTWRRRRRS-----ISRARQVELLLVADSMARLYRGLQHYLLTLASTAN 299
QY 251 RLYRHSILNPNIVVVKVLLLRDSDGPKYTGNAALTLRNFCAMQKKLNKVSQKHPEY 310
Db 300 RLYSHASIEHRLAVKVVVLDGDKSLEYSKNAATTLKNFCWKQHQNOLGDDHDEHY 359
QY 311 DTAILTRDLCGATTCDTLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMP 370
Db 360 DAAILFTREDLCGHHSCTDLGMADVGTICSPERSCAVIEDDGLHAAFTVAHEIGHLLGLS 419

```

```
QY 371 HDNVKVEEFGKLRANHMSPFLIOIDRANPWSACSAIIITDFLDGSHGDCLLDQPSKP 430
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 420 HDOSKCEETFGTEGDKRLMSSILTSIDASKPWSKCTSATIETFDLGHGNCLLDPRKQ 479
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 431 ISLPEDLPAGSYTLISOQCELAFLGKSGKPCPYMOYCTKLMCTGAKAGOMVQOTRHPFWADG 490
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 480 ILGPEELPGQTYDATQCNLTFGPEYSVCPGMDVCARLWCAVVROGMVCLTKKLPAVEG 539
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 491 TSGEGKCLKGACVER-----HNLNKHVRDGSWAKWDYPGCSRFGCGGVOLARRQCTN 545
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 540 TPCGKGRICLQKCVDKTKKKYISTSSH---GNMWSGWSGQCSRCGGGVGFAYRHCCN 596
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 546 PTPANGKYCEGVRKYRKNLPCPSSASGKSFREOCEAFNGYNHSTNRLTLAVAWVP 605
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 597 PAPRNGRYCTGKRAIYRSCSLMPCP--PNGKSFREOCEAKNGYQSDAKGVKTFVFWVP 654
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 606 KYGVSVPDRCKLICRANGTGYVVLAPKVVDTLCSPDSTSVCVQKCIKAGCCDNLGS 665
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 655 KYAGVLPADVCKLTCRAKGTGYVWFSPKVTDTGTECRPYSNSVCRGKCVRTGCDGIIGS 714
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 666 KKRFDKCGVCGGDNKCKKVTG 687
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 715 KLOYDKCGVCGGDNSSCTKIVG 736
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
US-09-130-491-13
; Sequence 13, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; CURRENT FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-130-491-13

Query Match 41.0%; Score 1536; DB 4; Length 608;
Best Local Similarity 63.6%; Pred. No. 1.9e-120;
Matches 269; Conservative 57; Mismatches 91; Indels 6; Gaps 4;

QY 269 VLLLRDRSGPKVTGNAALTLRNFCANQKLNKYSKDHPEYWDTAIFLTRODLGATTC 328
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 ILVIHDEQKGEVTSNAALTLRNFCNWKQHNPPSDRDAEYDIAIFLTRODLGSGTCD 63
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 329 TLGMADVGTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEVFGKLRANH 388
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 TLGMADVGTGMDPKRSCSVIEDDGLQAFTTAHELGHVFNMPHDNAKQACASLNGVND 123
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 389 MMSPTLIOIDRANPWSACSAIIITDFLDGSHGDCLLDQPSKPISLPDLPCASYTL 448
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 124 MNASMLSLDHSQWSPSCSAYMITSFLDNHGECLMDKPQNPIQLPGLDPTSDANRQC 183
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 449 ELAFGVSQKPCP-YMQYCTKLMCTGAKAGOMVQOTRHPFWADGTSCGEGKCLKGACVER 507
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 184 QTFDESKHCPDAASTCTSLWCTGTSGLVLCQTKKHPWADGTSCGEGKWCINGKCVNK 243
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 508 HNLNKH---RVDGSWAKWDYPGCSRFGCGGVOLARRQCTNPTPANGKYCEGVFVYKRS 564
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 244 TD-RKHFDTPFHGSGMGMGWDGCSRTCGGGVQVYTMRECDNPVPKNGGKYCEGRVYRS 302
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 565 CNLEPCSSASGKSFREOCEAFNGYNHSTNRLTLAVAWVPKYSGVSPDRCKLICRANG 624
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 303 CNLEDCPDN-NGKTFREOCEAHNEFSKASFGSGPAVENIWKTAGVSPKDRCKLICQAKG 361
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 625 TGYFVLAPKVVDTLCSPDSTSVCVQKCIKAGCCDNLGSKKRFRDKCGVCGGDNKCKK 684
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 362 IGYFVLQPKVVDTGTPCSPDSTSVCVQOQCVKAGCDRIIDSKKFKDKCGVCGGNGSTCKK 421
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 685 VTG 687
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 422 ISG 424
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7
US-09-369-364A-13
; Sequence 13, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1882
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-9
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (468)
; OTHER INFORMATION: xaa = C
; NAME/KEY: MOD_RES
; LOCATION: (521)
; OTHER INFORMATION: xaa = Y
US-09-369-364A-13

Query Match 38.2%; Score 1432; DB 4; Length 1882;
Best Local Similarity 42.1%; Pred. No. 4.6e-111;
Matches 293; Conservative 98; Mismatches 235; Indels 70; Gaps 15;

QY 22 EREVVPVIRLDP-----DINGRRYVW-----RGPEDSGDOGLFOITAFQ 61
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 49 EYEIVSPIRVNALGEPFTNVHFKTRRSINSAITDPWPAFASSSSSSTSSQAHYRLSAFG 108
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 EDYFLHLPDAQFLAPAFSTEHLGVP-----LQGLTGSSDLRRCFYSGDVNAEPDSFAAV 117
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 109 QQLFLNLTANAGFIAPLFTVLLGTGPGVNOTKPYSEEAELKHCFVK-----155
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 118 SLGGGLRGAFYRGAEYVVISPLPNASAPAAQNSQGAHLQRRGVGPGSGDPTSRGVA 177
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 156 LRCQYQLRAHG-----RHQPLLNRHKNRHSKDKKTRARKWGERINLAGDV 202
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 178 SGNNPAI-LRALDPYKPRAGFCESRRRSRGRKRVFVPIRYVETLVVADESVMKEHGA 236
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 203 AALNSGLATEAFSAYGNKTDNTRKTHRT---KFELSPREVEVLVVADNRWVSYHGE 259
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 237 DLEHYLLTLTAARLYRHPSTILNPINVVVLLLRDRSGPKVTGNAALIRNFCAWQ 296
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 260 NLQHYILTLMSIVASIKDPSIGNLINIVNLIVHNEQDGPSSISFNAQTTLKNFCQWQ 319
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 297 KKLNVKSKHPEYWDTAIFLTRODLGCA-TTCDTLGMADVGTGMDPKRSCSVIEDDGLPS 355
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 320 HSNSPGGIHH---DTAVLLTRQDICRAHDKCDTLGLAEGLTICDPRSCSISDSGLST 375
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 356 AFTAHGELGHVFNMPHDNVKVCVEVFGKLRANHMSPFLIOIDRANPWSACSAIIITDFL 415
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 376 AFTIAHELGHVFNMPHDNNKKEE-GVRSQPQHVMAPTLNFYTNPNWMSKSRKYITEFL 434
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 416 DSGHGDCLLDQP-SKPISLPDLPGASYTLISOQCELAFLGVSQKPCPYMOCYCTKLWC---T 471
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 435 DTGCGCLLNEPSPYPLPVLPGILLYNVNKOXELIFGPGSQVCPYMQCRRLWCNNVN 494
QY 472 GRAKQOMVQTRHFPWADTSCGEGKCLKLGACVERHNLNKRVDGSGNAKWDPPGPCSRT 531
Db 495 GVHKG---CRTQHTPWADTECEPGKHCKXGFCVPK-EMDVPTVDGSGWSWSPFTCSRT 550
QY 532 CGGGVQLARROCTNTPANGGKYGCVGVKVRKSCNLEPCPSASGKSPREOCEAFNGYN 591
Db 551 CGGGIKTAIRECNRPEPKNGGKYGVRMRKFKSCNTEPCLKQK---RDFRDEQCAHFDGKH 608
QY 592 HSTNRLTLAVAWPKYSGVSPDRCKLICKRANGTGYEVVLAPKVVVDGTLCSPDSTSVQVQ 651
Db 609 FNINGLLPNVRWPKYSGILMKDRCLFCRVAGNTAYTQLDRVVDGTGPGCGDNDICVQ 668
QY 652 GKICRAGDGNLGSKKRFDKCGVCGGDNKSKKVTG 687
Db 669 GLCROAGCDHVLNSKARRDKCGVCGGDNSSCKTVAG 704

RESULT 8
US-09-369-364A-15
; Sequence 15, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hurskainen, Tiina L.
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 15
; LENGTH: 874
; TYPE: PRT
; ORGANISM: Mus musculus ADAMTS-9
US-09-369-364A-15
```

```
Query Match 37.2%; Score 1394.5; DB 4; Length 874;
Best Local Similarity 44.6%; Pred. No. 2.2e-108;
Matches 271; Conservative 86; Mismatches 202; Indels 49; Gaps 12;

QY 115 AAVSLCGGLGAFAGYGEYVISLPLNAPAAQNSGGAHLORRGVPGSGDPTSRC 174
Db 3 AVISLCGMMGTFRSHDGDYFIEPLQSVDEQEEQNKPHIYRHSTPQREPSTGKHAC 62
QY 175 GVASGNP-----AILRA-----LDPKPRAGGESRR 205
Db 63 ATSELKNSHKDKRIRMKRRKRNSLADVALLKSLGATKVLGYSNQ-----NNTDRW 119
QY 206 RSGRAKRVSTPRVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSLNPNIV 265
Db 120 NKRTRKPLSYSPREVMVADHRMVLVHGANLQHYILTMSIVASYKDSIGNLNIV 179
QY 266 VVKVLLLRDRSGPKVTGNAALITRNCAWQKLNKYSV-KHPYWDTAILFTQDLCGA 324
Db 180 IVNLVVIHNEQGYINFAQTTLNFCQWQHSKNYLGGIQH-----DTAVLVTRDICRA 235
QY 325 -TTCDTGMADVGMCDPKRSCSVIEDGLPSAFTTAHELGHVFNMPHDNVKVEEYFGK 383
Db 236 QDKCDTLGLAELGICDPYRSCSISESGLSAFTTAHELGHVFNMPHDNSNCKEE-GV 294
QY 384 LRANHMSPTLIQIDRANPWSACSAIITDFLDGSHGDCILLDQ-SKPISLPDLPGASY 442
Db 295 KSPQHVMAPTLFNYTNPMWMSKSRKVIETFDLTGYGECLLNEPASRTYPLPSQLPCLLY 354
QY 443 TLSQOCELAFVGSKPCPYMQYCTKLWCT---GKAKQMVQCTRHFPWADTSCGEGKLC 499
Db 355 NVNKOCELIFGPGSQVCPYMQCRRLWCNNVDGAHKG---CKTQHTPWADTCEPEGKHC 411
QY 500 LKGAACVERHNLNKRVDGSGNAKWDPPGPCSRTCGGVLARRCTNTPANGGKYCEGV 559
```

```
Db 412 KFGFCVPK-EMEGPAIDGSMGWSHFGTCSRTCGGKIKTAIRECNRPEPKNGKYGVR 470
QY 560 VKYBSCNLEPCPSASGKSPREOCEAFNGYNHSTNRLTLAVAWPKYSGVSPDRCKLI 619
Db 471 MKFKSCNTEPCMKQK---RDFREOCEAHFDGKHFNINGLLPSPVRFKYSILMKDRCKLF 528
QY 620 CRANGTGYFYVLAPKVVVDGTLCSPDSTSVQVGKCIKAGCDGNLGSKKRFDKCGVCGGD 679
Db 529 CRVAGNTAYQLDRVVDGTGPGCGDNDICVQGLCROAGCDHILNSKVRDKCGICGGDN 588
QY 680 KSCKKVTG 687
Db 589 SSCKTVAG 596

RESULT 9
US-09-369-364A-7
; Sequence 7, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hurskainen, Tiina L.
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
US-09-369-364A-7
```

```
Query Match 28.8%; Score 1077; DB 4; Length 997;
Best Local Similarity 38.4%; Pred. No. 1.1e-81;
Matches 284; Conservative 82; Mismatches 266; Indels 108; Gaps 32;

QY 1 MLLGLITLAFAGTAGGPEPER---EVVVPRLDP-----DINGRR- 39
Db 18 LLLLCALAPGAPGAPGPRATEGRAALDIVHVRVDAGGSFLSYELMPRALRKRDVSRVD 77
QY 40 ---YY---WRGPRDSGDQGLIFQITAFQEDFYHLTPDAQFLAPAP--STEHLGVPQLGL 91
Db 78 APAFYELQYGRE-----LRFNLITANQ-----HL-----LAPGVSTRRG-----GL 116
QY 92 TGSSDLR---RCFYSGDVNAEPD---SFAAVSLCGGLRGAFGYRGAEYVISPLPNASA 144
Db 117 --GRAHIRAHTPACHLLGEVQ-DPELEGGGLAAISACDGLKGVFQLSNEYFIEPLD--SA 171
QY 145 PAAQRNSOGAHLLOQRGVP-----GGPSGDPTSRCVAGSGWNPAILRALDPYKPRRAG 199
Db 172 PAPGHAQ--PHVYKROAPERLAQRGDSAP-STCGV-----QYVPELESRRRW 219
QY 200 ESRSRRSRAKRF-----VSIPRYVETLVVADESVMKFGH-ADLEHYLLTLATAARLYR 254
Db 220 EQRQQRRLRLRHLRHSYKSKWCETLVVADAKWVEYHQPOVESYVLTINWVAGLPH 279
QY 255 HPSILNPINIVVVKVLLLRDRSGPKVTGNAALITRNCAWQKLNKVKSDKHPEYWDTAI 314
Db 280 DFIGNPIHITIVRLVLEDEEDLKITHADNTLKSFCWKQKSNMKMDAHLPHHDTAI 339
QY 315 LFTRODLCCATT--CDTLGMADVGMCDPKRSCSVIEDGLPSAFTTAHELGHVFNMPHD 372
Db 340 LTRKDLCAAMNRPCTLGLSHVAGMCQPHRSCSINEDTGLPLAFTVAHELGHSGFIQHD 399
QY 373 -NVKCVCEEYFGKLRANHMSPTLIQIDRANPWSACSAIITDFLDGSHGDCILLDQSPK 430
Db 400 GSGNDCFPV-GK--RFFIMSPQLLDAAAPLTWRSRQRVITITFDLGRGWLGLDDPPAKDI 456
QY 431 ISLPDLPGASYTLSQOCELAFVGSKPCPYM-QYCTKLWCTGKAKGMVQCTRHFPWAD 489
```

```
Db 457 IDPSPVPGVLYDYSHQCRQLQYGAYSAFCEDMDNVCHTLWCSVGT----TCHSKLDAAYD 512
;
QY 490 GTSCEGEGCLKLGACVERHNLKRVHVDGSAKWADPYGPCSRTCGGVLARRQCTNPTTPA 549
;
Db 513 GTRCGENKWSGECVP--VGFPRPAVDGGSGWSAWISCRSGMGVQSAERQCTQPTPK 571
;
QY 550 NGKVCCEGVRYKRSNLEPCSPSSAGSKSFREOCEAFNGYNHSTNRLTLAVAWVPKYS 609
;
Db 572 YKGRVCVGERKFRCLNLOACP--AGRPSFRHVQCSHFDAWLYKGQHLT-----WVPVND 625
;
QY 610 VSPRDKCLICRANGTYFYVVLAPKVVDGTLCS--DSTSVCVQGGKICAGCDGNLGSKK 667
;
Db 626 VNP---CELHCRPANEXYFAKRLRDACVDGTFCYQVRASRLDLCINGICKNVGCDFEIDSGA 682
;
QY 668 RFDKCGVCGDNKSKKVTG 687
;
Db 683 MEDRCGVCHGNGSTCHIVSG 702
;
RESULT 10
US-09-130-491-16
; Sequence 16, Application US/09130491
; Patent No. 6416974
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/09/130,491
; EARLIER FILING DATE: 1998-08-07
; EARLIER APPLICATION NUMBER: US 60/058,108
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: US 60/054,961
; EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Rattus rattus
US-09-130-491-16
Query Match 28.6%; Score 1072; DB 4; Length 551;
Best Local Similarity 58.7%; Pred. No. 1.2e-81;
Matches 185; Conservative 45; Mismatches 79; Indels 6; Gaps 4;
QY 377 CREVEFGKLRANHMSPPTLIQIDRANPWSACSAIITDFLDSGHGDCLLDQPSKPSISLPED 436
;
Db 1 CASLNGVSGDHLNLSLSDHSPQSPCSAYMVTSELDNGHGECLMDRKPQNPILPSPD 60
;
QY 437 LPGASYTLSSQCELAFLGVSCKPCP-YMQYCTKLWCTKAKGQMVQCTRHPFWADGTSCEG 495
;
Db 61 LPTGLYDANRQCQFTFGESKHCDFDAASTCTTLWCTGTSGLLYVCQTKHFPWADGTSCEG 120
;
QY 496 GKLCILKACVERHNLNKH---RVDGSAKWADPYGPCSRTCGGVLARRQCTNPTTPANG 552
;
Db 121 GKWCVSGKCVAKTDM-KHFAFPVHSGWGPWGPWGDSCSRTCGGGVQYTMRECDNVPVKNKG 179
;
QY 553 KYCEGVRYKRSNLEPCSPSSAGSKSFREOCEAFNGYNHSTNRLTLAVAWVPKYSVSP 612
;
Db 180 KYCEKRVRYKRSNLEDCPDN-NKTRFEEQCEAHNFESKASFGNETVETWTPRYAGVSP 238
;
QY 613 RDKCKLICRANGTYFYVVLAPKVVDGTLCSPDSTSVCVQGGKICAGCDGNLGSKKRFDK 672
;
Db 239 KDRCKLTCEAKGIGYFYVVLQPKVVDGTPCSPDSTSVCVQGGKICAGCDRIIDSKKRFDKC 298
;
QY 673 GVCGGDNKSKKVTG 687
;
Db 299 GVCGGNGSTCKRMMSG 313
;
RESULT 11
```

```
US-09-369-364A-22
; Sequence 22, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-5
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (99)
; OTHER INFORMATION: Xaa = M
US-09-369-364A-22
Query Match 26.8%; Score 1004; DB 4; Length 518;
Best Local Similarity 52.9%; Pred. No. 5.7e-76;
Matches 174; Conservative 42; Mismatches 103; Indels 10; Gaps 3;
QY 364 GHVFNNPDNKKVCEEVFGKLRANHMSPPTLIQIDRANPWSACSAIITDFLDSGHGDCLL 423
;
Db 1 GHLLGLSHDDSKFCEETFGSTEDKRLMSSILTSIDASKPWSKCTSATITFEFLDDGHGNC 60
;
QY 424 LDQSPKPSISLPEDLPGASYTLSSQCELAFLGVSCKPCPYMYCTKLWCTKAKGQMVQCTR 483
;
Db 61 LDLPKQILGPEELPGQTYDATQCNTLTFGPDYSVCPGXDVCARLWCAVVVRQGMVCLTK 120
;
QY 484 HFPWADGTSCEGKLCILKACVER-----HNLNKHVRDVGSAKWADPYGPCSRTCGGQV 538
;
Db 121 KLPAVEGTPCGKGRICLOGKCVDTKKKYSTSSH---GNWGSWGSQCSRSRCSGGVQF 177
;
QY 539 ARQCTNPTPANGKYCEGVRYKRSNLEPCSPSSAGSKSFREOCEAFNGYNHSTNRLT 598
;
Db 178 AYRHCNNPAPRNNGRYCTGKRAIYHSCSLMPCP--PNKSFREOCEAKNGYOSDAKGVK 235
;
QY 599 LAVAWVPKYSVSPRDKCLICRANGTYFYVVLAPKVVDGTLCSPDSTSVCVQGGKICAG 658
;
Db 236 TFEVWVPYAGVLPADVCLTKRAKGTGYVVFSPKVTDTGTECRPYNSVCVRGKCVRTG 295
;
QY 659 CDGNLGSKKRFDKCGVCGDNKSKKVTG 687
;
Db 296 CDGIIGSKLQYDKGCVCGDNDSSCTKIYG 324
;
RESULT 12
US-09-491-522-5
; Sequence 5, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Collige, Alain
; APPLICANT: Lapierre, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
```

OPERATING SYSTEM: Windows
SOFTWARE: FastSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/491,522

Query Match	26.1%;	Score 976;	DB 4;	Length 1211;
Best Local Similarity	35.5%;	Pred. No. 4e-73;		
Matches 260;	Conservative	93;	Mismatches 282;	Indels 98; Gaps 31;

QY	43	RGPE	SGDQGLIFQITAFQEDFYHLHTLPDQAFIAPAFSTHGLVPIQLQGLTGSSDLRRCF	101							
Db	99	GGNEE	PGSLFNYTVTFVGRDLHLRLRPNARLVAPGATMEWQCE--KGTRVEPLLGSL	156							
QY	103	YSGDVN-	-AEPD	FAAVSLCGLRGAFGYRGAEYVISPINSA	PAQNSQG-AHLLOR	159					
Db	157	YGDV	AGLAEASS-VALS	CNDGLIIRMEEEFFTEPLEKGL--AAQBAE	QGRVHVYR	213					
QY	160	RGVPG	SGDPTSCRGV	ASGWNPAI	LRLADPKPRRAGF	GESRRSRRAKRFVSI	PIR	219			
Db	214	RPPTS	PLGGPOA--LDTG--ASL	DSLDSL	S-RALGVLEE	HANSSRRARRHAADDY	266				
QY	220	VE-TLV	VADESMVKF	GAD-LEHY	LTLTATARLYR	HPHSILNIPINIV	VKVLLLRDRS	277			
Db	267	NEVL	LGVDSDSVVQFHG	KEVQKYLTLN	INEYIHDESL	GAHINVLVIRILL	SYGKS	326			
QY	278	GPXV-TG	NAALTURNE	CAWKLNK	YSDKHPEYD	FAILFTRODLC	GAITCDTLGNADVG	336			
Db	327	MSLIE	IGNPSQSL	ENVCWAYLOQ	EPDTCG	DEYHDIAIFLT	RQDF-GPSGMO-GYAPVT	383			
QY	337	TMCD	PRKSCVIEDDGL	PSAFTAH	ELGHVFN	MHDNV-KVCE	EEVFGKLRAHNMSP	PTLI	395		
Db	384	GMCH	PYRSTL	NHEDG	FSFAFVVAHETG	HVLGMEHD	GQGNRCGD--EVR	LGSIMAPLVQ	440		
QY	396	QIDRAN	PWASAAIITD	FLDLSHG	DCILLDQ--SKP	SILBEDLP	GSATYTL	SQCEALF	452		
Db	441	AAFR	HPWHRCSQ	UESRYSLHS--YDCL	LDDPFAHDWP-ALPQ-L	FLG	LYSMNS	QCRFDF	496		
QY	453	GVSK	PCPYMQY---CTKL	WCCTGKAKG	QWVCOTRHF	PMADGTSCEG	KLCLAGACV----	505			
Db	497	GLGY	MCTAFR	FDPCQLWCS-HPDN	PYFCTKKGP	PLDGTW	ACPKHCFKHG	CHILWLP	555		
QY	506	---	ERIN	LANKRVDG	SWAKWDYPG	CSRTCCGGV	QLARRQCTNPT	PANGKYCEG	VRVY	562	
Db	556	DILK	R-----DGS	WGAWS	PFSGSC	RTCGTGVK	FRTRQCDN	PHNPANG	GRTC	SLGAYDF	607
QY	563	RSCN	LEPCPSS	ASGKSF	FREEQ	CEANFY-----NH	STRNLT	TLAV	MPVKY	GVSPDRK	615
Db	608	QUCS	RDCDP	SLUA--D	FREEQ	RQNDL	YFEH	GDAQHH-----WLP-	HEHR	DAKER	654

```
QY 590 YNHSTNRLTLAVANVPKSGVSPROCKKLCICRANGTGYFYVLAPKVVDTGLCSFDSVSVC 649
Db 663 RHYK-----WKP-YTQVEDQDLCKLYCIAEGDFDFSLNKNKVDGTPCSEDSRNVC 712
QY 650 VQKCIKACDGNLGSKRKPKCGVCGDNKSC 682
Db 713 IDGICERVGDNLGSDAEDVCGVCGNNSAC 745

RESULT 14
US-09-491-522-11
; Sequence 11, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Collige, Alain
; APPLICANT: Lapierre, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-491-522-11

Query Match 25.2%; Score 944.5; DB 4; Length 1205;
Best Local Similarity 33.5%; Pred. No. 1.7e-70;
Matches 256; Conservative 96; Mismatches 276; Indels 137; Gaps 32;

QY 1 MLLGILFLAPAGTAG-----GFPEREVVPIRL----- 32
Db 16 LLLLLLLPDLAALAAADPPGPGQGHGAERILAVPVRTDAQGLVSHVSAATAPAG 75
QY 33 -----PDINGRYYWRGPDSDQGLFQITAFQEDFYHLHTPDAQFLAPAFST 81
Db 76 VTRRAAPAQIPGLSG-----GSEDPGGLFYNYTVFGDHLRLRFNARLVAPGATV 129
QY 82 EHLGVLQGLTGGSDLRRCFYSGDV--NAEPDSPAAYSLCGGLRGAFYRGAEYVISPPL 140
Db 130 EQWQE--SGATRVPEPLTGLTCLYVDGVAGLAESSVALSNCGLAGLIRMBEEFFIEPLE 187
QY 141 NASAPAAQRNSQG--AHLLORRGVPGSGDPTSR-----CGVASGWNPAIRAL 188
```

```
Db 188 KGL--AAKEAQGRVHVYHRT-----TSRFFLGFGQALDTGISADSLDSLRAL 237
QY 189 DPYKRRAGFGESESRRRSGRAKRFVSPRY-VETLVVADESVMVKFHGAD-LEHYLLTLL 246
Db 238 -----GVLEERVNSRRMRRAADDDYNIIEVLGVDDSVVQFHGTEHVQKYLTLTM 289
QY 247 ATAARLYRHPISILNPINIVVVKVLLLRDROSGPKV-TGNAALTLRNFCWAKOKLKNVSDK 305
Db 290 NIVNEIYHDESLGAHINVLVRIILLSYGKSMLEIGNFSQLENVCRWAYLOQKPDTD 349
QY 306 HPEYNDTAILFTRODLGCAITCDTLGMADYGTMDCKPKRSCSVIEDDGLPSAFTAHELGH 365
Db 350 HDEYHDHAIFLTRQDF-GPSGMQ--GYAPVTGMCHPVRSCTLNHEDGFSFAFVVAHETGH 406
QY 366 VENMPHDNV-KVCEEVFGKLRANHMSPTLIQIDRANPWSACSAAITDPLDSGHGDCLL 424
Db 407 VLGMHDGQGNRCGD---EVLGSIAPLVQAAAFHRFHWRSRCSQOELSRYLHS--YDCLR 461
QY 425 DQP---SKPISLPEDLPASVYLLSQCELAFGVGSKPCPYMQY---CTKLWCTGKAKGOM 478
Db 462 DDPFTHDWP-ALPO-LPGLHYSMNEQCRDFDGLGYMMCTAFRTFDPCKQLWCS-HPDNPY 518
QY 479 VCQTRH-FPWADGTSCGEGKCLAKGAC-----VERHNLNKHRVDGSKNAKWDPPYPCSR 530
Db 519 FCKTKKGFP-LDGTWCAPGKHCFKGCHTWTFTDLKR-----DGNMANSFPFGCSR 569
QY 531 TCGGVOLARROCTNPTPANGKYCEGVVRKYRSCNLEPCSPSSAGSKSPREECEAFNGY 590
Db 570 TCGTGVRKTRQCDNPHANGRTCSGLAYDFOLCNSQDCPDALA--DFREEQCRWDLY 627
QY 591 -----NHSTNRLTLAVANVPKSGVSPRDKCKLCICRANGTGYFYVLAPKVVDTGLCS- 642
Db 628 FEHGDAQHH-----WLP-HEHRDAKERCHLYCESKETGEVYSMKRMVMDGTRCSY 676
QY 643 PDSTSVCOVKICIRAGDGNLGSKKRFDKGVCGGDNKSKCKVTG 687
Db 677 KDAFSLVGRDCRVKVGCDGVIGSSKQEDKGVCGGDNHCKVKVG 721

RESULT 15
US-09-369-364A-17
; Sequence 17, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-10
US-09-369-364A-17

Query Match 24.9%; Score 930.5; DB 4; Length 1081;
Best Local Similarity 35.3%; Pred. No. 2.2e-69;
Matches 232; Conservative 82; Mismatches 276; Indels 67; Gaps 22;

QY 64 FYHLHTPDAQFLAPAFSTEHGLVPLQGLTGGSDLRRCFYSGDVNAE-PDSFAAVSLCGG 122
Db 65 FLLNTRSSRLAQRVSVETW--TREGIAWRARPHCLYAGHLQGAQSSHVVAITCGG 122
QY 123 LRGAFGYRGAEYVISPPLNAPAAQRNSQAHLLQRGVPGSGDPTSRCCV--ASGW 180
Db 123 LHGLIVADEEYLIETPLHGGPKGSRSPESGPHCVYKRSLRHPHLD--TACGVREKWP 180
QY 181 --NPAIRALDPYKPRRAGFGESESRRRSGRAKRFVSIPIRYVETLVVADESVMVKFHG-AD 237
```



```

Db 181 KGRPWMLTKLP--PPARPLGNETERGPG-LKRSYSRERYVETMDVADKMMVAYHGRD 237
QY 238 LEHYLLILLATAARLYRHPSTILNPINIVVVKVLLLRDRSGPKVTGNAALTLRNFCWOK 297
Db 238 VEQYVLAIMNTIVAKLFDQSLGSTVNILVTRILLTEDQPTLEITHAGKSLDSECKWOK 297
QY 298 KL-----NKVSKHPEYWDTAILFTRODLC--GATTCDTLGMADVGTMCDPKRSCSVI 348
Db 298 SIVNHSHGHNAPENGVAHDTAVLITRYDICIYKNKPCGTGLAR-WAECYSAREAAAS 356
QY 349 EDDGLPSAFTTAHELGHVFNPHDNVKEVEFG-----KLRANHMSPTLIQIDRANP-- 402
Db 357 MRTLAAATSVHHCHEIGHTFGMNHGVDGNSCGARGQDPAPKLMAAHITMKT-----NPFV 409
QY 403 WSACSAALITDFLDSHGDCILLDOPSK-PISLPEDLPGASYTILSQOCELAFAFGVSKPCPY 461
Db 410 WSSCNRDYITSFDSGLGLCLNRPQDFVYPTVAPGQAYDADAEOCRFQHGKVSQCKY 469
QY 462 MOYCTKLWCTGKAGQWQVQTRHPHPWADGTSCEGKL-----CLKGACVERHNLNKHVRVG 517
Db 470 GEVCSLWCLSKSNR---CITNSIPAAEGTICQHTTIDKGWCYKRVCP-FGSRPEGVDG 525
QY 518 SWAKWDYPCGCSRTCGGVOLARQCTNPTPANGGKYCEGVRYKYRSCNLEPCPSSASGK 577
Db 526 ANGPTWPGDCSRTCGGVSSSRHCDSPRTGGKYCLGERRRHRSCTNTDDCPPGS--Q 583
QY 578 SFREQCEAFNG-----YNHSTNRLTLAVAWVPKYSGVSPRDKCKLICRANGTYFYV 630
Db 584 DPREVQCAEFDSIPFRGKFKWKTYR-----GGGVK---ACSLTSLAEGFNFYTE 630
QY 631 LAPKVVDTGLCPDSTSVCVQCKIKAGCDGNLGSKKRFDKCGVCGDGNKSKKVTG 687
Db 631 RAAAVVDGTPCRPDTVICVSGECKHVGCORVLGSLREDKCRVCGDGSACETIEG 687

```

Search completed: April 29, 2003, 17:11:55
Job time : 15.1439 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:12:21 ; Search time 15.2713 seconds
(without alignments)
3604.758 Million cell updates/sec

Title: US-10-009-332-1_COPY_1_687

Perfect score: 3744

Sequence: 1 MLLGILTTLAFAGRTAGGFE.....RFDKCGVCGDNKCKKVTG 687

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/FCI_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/FCIUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3736	99.8	950	10	US-09-965-631-4
2	3707.5	99.0	823	9	US-10-163-316-2
3	1960.5	52.4	967	12	US-10-105-929-2
4	1947.5	52.0	950	10	US-09-321-387B-4
5	1947.5	52.0	968	9	US-10-163-316-7
6	1913	51.1	367	10	US-09-965-631-6
7	1751	46.8	727	9	US-10-097-597-1
8	1751	46.8	727	9	US-10-097-580-1
9	1738	46.4	727	10	US-09-445-023A-1
10	1738	46.4	727	9	US-10-097-597-12
11	1738	46.4	727	9	US-10-097-580-12
12	1738	46.4	727	10	US-09-445-023A-12
13	1688.5	45.1	837	9	US-10-174-590-352
14	1688.5	45.1	837	9	US-10-176-758-352
15	1688.5	45.1	837	9	US-10-175-737-352
16	1688.5	45.1	837	9	US-10-173-706-352
17	1688.5	45.1	837	9	US-10-175-738-352
18	1688.5	45.1	837	9	US-10-175-752-352
19	1688.5	45.1	837	9	US-10-176-482-352

20	1688.5	45.1	837	9	US-10-176-757-352	Sequence 352, App
21	1688.5	45.1	837	9	US-10-176-913-352	Sequence 352, App
22	1688.5	45.1	837	9	US-10-180-552-352	Sequence 352, App
23	1688.5	45.1	837	9	US-10-180-557-352	Sequence 352, App
24	1688.5	45.1	837	9	US-10-173-700-352	Sequence 352, App
25	1688.5	45.1	837	9	US-10-174-572-352	Sequence 352, App
26	1688.5	45.1	837	9	US-10-174-579-352	Sequence 352, App
27	1688.5	45.1	837	9	US-10-174-582-352	Sequence 352, App
28	1688.5	45.1	837	9	US-10-174-588-352	Sequence 352, App
29	1688.5	45.1	837	9	US-10-175-739-352	Sequence 352, App
30	1688.5	45.1	837	9	US-10-175-740-352	Sequence 352, App
31	1688.5	45.1	837	9	US-10-175-743-352	Sequence 352, App
32	1688.5	45.1	837	9	US-10-176-488-352	Sequence 352, App
33	1688.5	45.1	837	9	US-10-176-492-352	Sequence 352, App
34	1688.5	45.1	837	9	US-10-176-747-352	Sequence 352, App
35	1688.5	45.1	837	9	US-10-176-750-352	Sequence 352, App
36	1688.5	45.1	837	9	US-10-176-985-352	Sequence 352, App
37	1688.5	45.1	837	9	US-10-176-987-352	Sequence 352, App
38	1688.5	45.1	837	9	US-10-176-991-352	Sequence 352, App
39	1688.5	45.1	837	9	US-10-176-992-352	Sequence 352, App
40	1688.5	45.1	837	9	US-10-176-993-352	Sequence 352, App
41	1688.5	45.1	837	9	US-10-184-658-352	Sequence 352, App
42	1688.5	45.1	837	9	US-10-173-695-352	Sequence 352, App
43	1688.5	45.1	837	9	US-10-173-697-352	Sequence 352, App
44	1688.5	45.1	837	9	US-10-173-705-352	Sequence 352, App
45	1688.5	45.1	837	9	US-10-174-576-352	Sequence 352, App

ALIGNMENTS

RESULT 1
US-09-965-631-4
; Sequence 4, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Enco
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965.631
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 950
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-965-631-4

Query Match	99.8%	Score 3736;	DB 10;	Length 950;
Best Local Similarity	99.9%	Pred. No. 1.8e-276;		
Matches 686;	Conservative	0; Mismatches 1;	Indels 0;	Gaps 0;
Qy	1	MLLGILTTLAFAGRTAGGFEPEPEVVPVIRLDPDINGRRYWRGPDGQGLIFQITAF	60	
Db	1	MLLGILTTLAFAGRTAGGFEPEPEVVPVIRLDPDINGRRYWRGPDGQGLIFQITAF	60	
Qy	61	QEDFYHLTPDAQFLAFSTHGLVPLQGLTGGSSDLRRCFYSGDVNAEPDFAAVSLC	120	
Db	61	QEDFYHLTPDAQFLAFSTHGLVPLQGLTGGSSDLRRCFYSGDVNAEPDFAAVSLC	120	
Qy	121	GGLRGAFYRGAEVVISPLNAPAAQNSQGAHLQRGVPDGPSTGDDTSCRGVASGW	180	
Db	121	GGLRGAFYRGAEVVISPLNAPAAQNSQGAHLQRGVPDGPSTGDDTSCRGVASGW	180	
Qy	181	NPAILRALDPYKPRRAGFGESRRSRRAKRFVSIPIRYVETLVVADESVMKFGADLEH	240	
Db	181	NPAILRALDPYKPRRAGFGESRRSRRAKRFVSIPIRYVETLVVADESVMKFGADLEH	240	
Qy	241	YLTLTATAARLYRHPILLPINIVVVKVLLLRDRDSGPKVTGNALTLRNCAMOKKLN	300	

```

Db      241  YLUTLTAARLYRHPSILNPINIVVVKVLLLRDRSGPKVTGNAALTLRNFCAWQKLN 300
      |||
QY      301  KVS DKHPEYWDTAILEFTRODLGCAATCDTLGMADYGTMCDPKRSCSVIEDGLPSAFTTA 360
      |||
Db      301  KVS DKHPEYWDTAILEFTRODLGCAATCDTLGMADYGTMCDPKRSCSVIEDGLPSAFTTA 360
      |||
QY      361  HELGHVFNPHDNVNVKEEVFKGLRANHMSPTLQIDRANPWSACSAIITDFLDLSGHG 420
      |||
Db      361  HELGHVFNPHDNVNVKEEVFKGLRANHMSPTLQIDRANPWSACSAIITDFLDLSGHG 420
      |||
QY      421  DCLLDPSKPISLPELPGASYTLSQOCELAGVGSKPCPYMQYCTKLWCTGKAKGQWVC 480
      |||
Db      421  DCLLDPSKPISLPELPGASYTLSQOCELAGVGSKPCPYMQYCTKLWCTGKAKGQWVC 480
      |||
QY      481  QTRHFPWADGTSCEGKLCILKGACVERHNLNKHRYDGSWAKWDPYGCSRTCGGVQQLAR 540
      |||
Db      481  QTRHFPWADGTSCEGKLCILKGACVERHNLNKHRYDGSWAKWDPYGCSRTCGGVQQLAR 540
      |||
QY      541  RQCTNPTPANGKYCEGVRVVKYRSCNLEPCPPSSASGKSPFEEQCEAFNGYNHSTNRLTLA 600
      |||
Db      541  RQCTNPTPANGKYCEGVRVVKYRSCNLEPCPPSSASGKSPFEEQCEAFNGYNHSTNRLTLA 600
      |||
QY      601  VAWVPKYSGVSPRDKCKLICRANGTCGYFYVLA PKVVDGTLCSPDSTSVCVQCKICAKGCD 660
      |||
Db      601  VAWVPKYSGVSPRDKCKLICRANGTCGYFYVLA PKVVDGTLCSPDSTSVCVQCKICAKGCD 660
      |||

QY      661  GNLSKRRFRDKCGVCGGDNKSKCKVTG 687
      |||
Db      661  GNLSKRRFRDKCGVCGGDNKSKCKVTG 687
      |||

RESULT 2
US-10-163-316-2
; Sequence 2, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MPI01-025PIRNM
; CURRENT APPLICATION NUMBER: US/10/163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-316-2

```

```

241 YLTLTLLAARLYRHPISILNPINIVVVKVLLLRDRDGPVKVTGNAALTLRNFAWOKKL 300
    |||
    |||
    |||
241 YLTLTLLAARLYRHPISILNPINIVVVKVLLLRDRDGPVKVTGNAALTLRNFAWOKKL 300
    |||
    |||
    |||
301 KVS DKHPEWDTAILFTRODLGATCTDGLMADVGTMCDPKRS CVIEDDGLPSAFTTA 360
    |||
    |||
    |||
301 KVS DKHPEWDTAILFTRODLGATCTDGLMADVGTMCDPKRS CVIEDDGLPSAFTTA 360
    |||
    |||
    |||
361 HELGHVFNMPHDNVKVEEVFKLKRANHMMSPTLIQIDRANPWSACSAAIITDPLDSGHG 420
    |||
    |||
    |||
361 HELGHVFNMPHDNVKVEEVFKLKRANHMMSPTLIQIDRANPWSACSAAIITDPLDSGHG 420
    |||
    |||
    |||
421 DCLLDQPSKPISLPELPGASVTLT SQOCELAFGVGSKPCPYMQVCTKLWCTGKAKGOMVC 480
    |||
    |||
    |||
421 DCLLDQPSKPISLPELPGASVTLT SQOCELAFGVGSKPCPYMQVCTKLWCTGKAKGOMVC 480
    |||
    |||
    |||
481 QTRHFPAW DGTSCGEGKLC LGACVERHNLNKHR-----VDG 517
    |||
    |||
    |||
481 QTRHFPAW DGTSCGEGKLC LGACVERHNLNKHRPTDIISPQKLLRLPNGLHTTQVDG 540
    |||
    |||
    |||
518 SWAKWDYPGCSRTC GGGVQLARRQCTNTPANGSKYCEGVVRKYRSCNLEPCCPSSASGK 577
    |||
    |||
    |||
541 SWAKWDYPGCSRTC GGGVQLARRQCTNTPANGSKYCEGVVRKYRSCNLEPCCPSSASGK 600
    |||
    |||
    |||
578 SFREQCEAFNGYNHSTNRLT LAVAWPKYSGVSPRDKCLICRANGGYFYVLAPKVV D 637
    |||
    |||
    |||
601 SFREQCEAFNGYNHSTNRLT LAVAWPKYSGVSPRDKCLICRANGGYFYVLAPKVV D 660
    |||
    |||
    |||
638 GTLCS PDSTSVCVQGCKICAGCDNGLSGKKRFDKCGVCGDNKSKCKVTG 687
    |||
    |||
    |||
661 GTLCS PDSTSVCVQGCKICAGCDNGLSGKKRFDKCGVCGDNKSKCKVTG 710
    |||
    |||
    |||

RESULT 3
US-10-105-929-2
; Sequence 2, Application us/10105929
; Patent No. US20020137142A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 0940/041001
; CURRENT APPLICATION NUMBER: US/10/105,929
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/130,491
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-105-929-2

```

QY 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPKPRRAGGESRRRS 207
Db 197 TCGVDDPRPTGKAETEDDEGEDEGPQWS-----PODPALQGVGP-TGTGS 247
QY 208 GRAKRFVSIPRYVETLVVADESVMYKFGHADLEHYLLTLTATAARLYRHPSTILNPINWV 267
Db 248 IRKKRFVSHRYVETMLVADQSMAEFGHGLKHVLLTLFSAARLYKHPSTIRNSVSLVV 307
QY 268 KVLRLDRDSDGPKVTGNAALTLRNFCAWKKLNKVSOKHPYWDYTAIFTRQDLCGATTC 327
Db 308 KILVTHDSQKGPVETSNAAALTLRNFCAWKKLNKVSOKHPYWDYTAIFTRQDLCGATTC 367
QY 328 DTLMADYGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNLVNVKVEEVEGKLRAN 387
Db 368 DTLMADYGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNLVNVKVEEVEGKLRAN 427
QY 388 HNSPTLIQIDRANPWSACSAIITDFTDLSGHGCLLDQPSKPSLSPEDLPASVYTLSSQ 447
Db 428 HMASMLNLDHSPWSPCSAYMITTFDNGHGECLMDKPNPQLPDLPGTSDANRQ 487
QY 448 CELAFGVGSKPCP--YMOYCTKLWCTGKAKGOMVQOTRHFPAWADTSGEGKCLKLGACVE 506
Db 488 COFTGEDSKHCPDAASTCTTLWCTGSGVLCVQTHFPAWADTSGEGKCLKLGACVE 547
QY 507 RHNLNKH---RVDGSAWAKWDPYGCRTCGGGVQLARRQCTNPTPANGKGYCEGVRYK 563
Db 548 KTD-RKHEDTTFHSGWGMWGPWDCSRCTCGGVQYTMRECDNPVPKNGKGYCEGVRYK 606
QY 564 SCNLEPCPSSASGSKFRECEAEANGYNHSTNRLTLAVAVVPKYSVSPDKCKLICRAN 623
Db 607 SCNLEPCPSSASGSKFRECEAEANGYNHSTNRLTLAVAVVPKYSVSPDKCKLICRAN 665
QY 624 GTGYFVYLAPKVDGTLCSPDSTSVCGVQKCIKAGCDNGLSKKRPDKCGVCGDNKSK 683
Db 666 GIGYFVYLAPKVDGTLCSPDSTSVCGVQKCIKAGCDNGLSKKRPDKCGVCGDNKSK 725
QY 684 KVTG 687
Db 726 KISG 729

RESULT 4
US-09-321-987B-4
; Sequence 4, Application US/09321987B
; Patent No. US20020102210A1
; GENERAL INFORMATION:
; APPLICANT: Kimble, Judith E
; TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
; FILE REFERENCE: 960296,95386
; CURRENT APPLICATION NUMBER: US/09/321,987B
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087,170
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/129,023
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Murine
US-09-321-987B-4

Query Match 52.0%; Score 1947.5; DB 10; Length 950;
Best Local Similarity 52.3%; Pred. No. 3.5e-140;
Matches 381; Conservative 95; Mismatches 176; Indels 77; Gaps 16;

QY 1 MLLGILTLAPAGTAGG--FEPEREVVPVIRLDPDINGRRYWRGP-EDSGDQGLIFOI 57
Db 20 LLLASITMLLCARGHGRTEEDLVLPSLE-----RAPGHDSTTRL--RL 66

QY 58 TATQDEPYLHLTPDAQFLAPAFSTEHGLVPLQGLTGS-----SDLRRCFYSGDVNA 109
Db 67 DAFGQQLHLKLPDSDGFLAPGFTLTQTV-----GRSPGSEAQHLDPDGLAHCFYSGTVNG 121
QY 110 EPDSFAVSLCGGLRGAFGYRGAEEYVSLPLPNAS---APAAQNSOGA---HLLQRRG 161
Db 122 DPGSAALSLCEGVRGAFYLOGEEFTIQAPGVATRELAPAVPEEESARPOFHILRRR- 180
QY 162 VPGPGSDPTSRGCVASGWNPAILRALDPYKPRRAGFESRSRRR----- 206
Db 181 -----RRSGGAKGCVMD-----DETLPSTDSPESQNTNRNQPVRDPTPDAGRP 226
QY 207 -----SGRAKRFVSIPRYVETLVVADESVMYKFGHADLEHYLLTLTATAARLYRHPSTILNP 262
Db 227 SGPGSIRKKRFVSHRYVETMLVADQSMADFHGSLGLKHVLLTLFSAARLYKHPSTIRNSI 286
QY 263 NIYVVVLLLRDRSDGPKVTGNAALTLRNFCAWKKLNKVSOKHPYWDYTAIFTRQDLC 322
Db 287 SLVVVKILVYEBQKGPVETSNAAALTLRNFCAWKKLNKVSOKHPYWDYTAIFTRQDLC 346
QY 323 GATTCDTLMADYGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNLVNVKVEEVEG 382
Db 347 GSHTCTDLMADYGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNLVNVKVEEVEG 406
QY 383 KLRANHMSPTLIQIDRANPWSACSAIITDFTDLSGHGCLLDQPSKPSLSPEDLPASVY 442
Db 407 VTGDSHLMASMLNLDHSPWSPCSAYMITTFDNGHGECLMDKPNPQLPDLPGTSDANRQ 466
QY 443 TUSQCELAFAFGVSKPCP--YMOYCTKLWCTGKAKGOMVQOTRHFPAWADTSGEGKCLK 501
Db 467 DANROCOFTFGBESKHCPCDAASTCTTLWCTGSGVLCVQTHFPAWADTSGEGKCLK 526
QY 502 GACVERHNLNKH---RVDGSAWAKWDPYGCRTCGGGVQLARRQCTNPTPANGKGYCEGV 558
Db 527 GKCYNKTD--KHFPATPVHSGWGMWGPWDCSRCTCGGVQYTMRECDNPVPKNGKGYCEG 585
QY 559 RVKRYSCNLEPCPSSASGSKFRECEAEANGYNHSTNRLTLAVAVVPKYSVSPDKCKL 618
Db 586 RVKRYSCNLEPCPSSASGSKFRECEAEANGYNHSTNRLTLAVAVVPKYSVSPDKCKL 644
QY 619 ICRANGTGYFVYLAPKVDGTLCSPDSTSVCGVQKCIKAGCDNGLSKKRPDKCGVCGGD 678
Db 645 TCEAKGIGYFVYLAPKVDGTLCSPDSTSVCGVQKCIKAGCDNGLSKKRPDKCGVCGGD 704
QY 679 NKSKKVTG 687
Db 705 GSTCKKMSG 713

RESULT 5
US-10-163-316-7
; Sequence 7, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Liebermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; FILE REFERENCE: MPI01-025P1RN
; CURRENT APPLICATION NUMBER: US/10/163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-163-316-7

Query Match 52.0%; Score 1947.5; DB 9; Length 968;
Best Local Similarity 52.3%; Pred. No. 3.6e-140;
Matches 381; Conservative 95; Mismatches 176; Indels 77; Gaps 16;

[illegible]

```

RESULT 6
US-09-965-631-6
; Sequence 6, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Fridgde, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Encoding
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 367
; TYPE: PRT
; ORGANISM: homo sapiens

```

[illegible]

```

RESULT 7
US-10-097-597-1
; Sequence 1, Application US/10097597
; Publication No. US20030022352A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiji
; APPLICANT: Hakezaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: pharmaceutical
; TITLE OF INVENTION: composition and method of immunologically analyzing human ADA
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-597-1

```

```

Query Match          46.8%; Score 1751; DB 9; Length 727;
Best Local Similarity 64.0%; Pred.No. 2.e-125;
Matches 309; Conservative 68; Mismatches 100; Indels    6; Gaps
                                4;

QY   209 RAKRFVSIPIRYVETILVVADSMVKFHGADLEHYLLTLTATAARLYRHPISLNPINIVVK 268
      |||||.....|||||..|||||...|||||..|||||..|||||..|||||..|||||..
Dh   q RKHFVFSSPRYYVTMLVADOSMAEERHSGSLKVIYLTLTFSVAAARLYKHYPHSIRNSLVVKK 68

```


Db 249 QFTFGESKHCDAASTCTTTLWCTSGGLLVQCFKHFPMADGTSCEGKWCYSGKCVNK 308
QY 508 HNLNKH---RVDSWAKWDPGPCSRCTCGGVLARRQCTNPTPANGKCYCEGVKRYKRS 564
Db 309 TDM-KHFATPVHGSWPGMGDCSRTCGGVQYTMRECDNPVKNGKCYCEGKRVYRS 367
QY 565 CNLEPCPSASGKSFRFEEQCEAFNGYNHSTNRLTLAVAWVPKYSGVSPDRCKLICRANG 624
Db 368 CNIEDCPDN-NGKTFREEQCEAHNEFSKASFGNEPTVETWPKYAGVSPDRCKLICRANG 426
QY 625 TGYFYVLAPKVVDTGLCSPDSTSVCVQGCICKAGCDGNLGSKKRFDKCGVCGGDNKSKCK 684
Db 427 IGYFFVLQPKVVDTGPCSPDSTSVCVQGCICKAGCDRIIDSKKKFKDKCGVCGGNGSTCKK 486
QY 685 VTG 687
Db 487 MSG 489

RESULT 12

US-09-445-023A-12

; Sequence 12, Application US/09445023A

; Patent No. US20020119167A1

; GENERAL INFORMATION:

; APPLICANT: Hirose, Kunitaka

; APPLICANT: Inoguchi, Eiji

; APPLICANT: Hakoziaki, Michinori

; APPLICANT: Ishioka, Keiko

; APPLICANT: Ishida, Yukako

; APPLICANT: Matsushima, Kouji

; APPLICANT: Kuno, Kouji

; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical

; FILE REFERENCE: Q57092

; CURRENT APPLICATION NUMBER: US/09/445,023A

; PRIOR FILING DATE: 1999-12-03

; PRIOR FILING DATE: 1997-06-03

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 12

; LENGTH: 727

; TYPE: PRT

; ORGANISM: Mus sp.

US-09-445-023A-12

Query Match 46.4%; Score 1738; DB 10; Length 727;
Best Local Similarity 63.4%; Pred. No. 2.3e-124;
Matches 306; Conservative 68; Mismatches 103; Indels 6; Gaps 4;

QY 209 RAKRFVSPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPNIVVVK 268

Db 9 RKRFVSPRYVETLVVADQSMADFHGSLGKHYLLTLFVAARLYRHPSIRNSISLVVVK 68

QY 269 VLLLRDRSGPKVTGNALTLRNFCWAKQKLNKYSKHPYWDTAIFLTRQDLCSGATCD 328

Db 69 ILVIEQKGPVTSNAALTLRNFCWQKHNSPSDRDPEHYDTAIFLTRQDLCSGHTCD 128

QY 329 TLGMADVTGCDPKRSCSVIEDDGLPSAFTTAHELGHVFNHNDNVKVCVEFGKLRANH 388

Db 129 TLGMADVTGCDPKRSCSVIEDDGLQAAFTTAHELGHVFNHNDNVKVCVEFGKLRANH 188

QY 389 MNSPTLIQIDRANPWSACSAIIIFDLSDGHDCLLDLPKSPISLPEDLPASVYTLSSQOC 448

Db 189 LMASMLSSLDHSPQSPSCSAVMTSFLDNGHGECLMDKPNQPKLSDLPGLTYLDNRQC 248

QY 449 ELAFGVGSKPCP-YMQVCTKLWCTGKAKQMVQOTRHFHPWADGTSCEGKLCCLKGACVER 507

Db 249 QFTFGESKHCDAASTCTTTLWCTSGGLLVQCFKHFPMADGTSCEGKWCYSGKCVNK 308

QY 508 HNLNKH---RVDSWAKWDPGPCSRCTCGGVLARRQCTNPTPANGKCYCEGVKRYKRS 564

Db 309 TDM-KHFATPVHGSWPGMGDCSRTCGGVQYTMRECDNPVKNGKCYCEGKRVYRS 367
QY 565 CNLEPCPSASGKSFRFEEQCEAFNGYNHSTNRLTLAVAWVPKYSGVSPDRCKLICRANG 624
Db 368 CNIEDCPDN-NGKTFREEQCEAHNEFSKASFGNEPTVETWPKYAGVSPDRCKLICRANG 426
QY 625 TGYFYVLAPKVVDTGLCSPDSTSVCVQGCICKAGCDGNLGSKKRFDKCGVCGGDNKSKCK 684
Db 427 IGYFFVLQPKVVDTGPCSPDSTSVCVQGCICKAGCDRIIDSKKKFKDKCGVCGGNGSTCKK 486
QY 685 VTG 687
Db 487 MSG 489

RESULT 13
US-10-174-590-352
; Sequence 352, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P343031C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; PRIOR FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-352

Query Match 45.1%; Score 1688.5; DB 9; Length 837;
Best Local Similarity 47.2%; Pred. No. 1.6e-120;
Matches 328; Conservative 103; Mismatches 215; Indels 49; Gaps 14;

QY 1 MLLGLTLTAFAGRTAGGPEPEREVVPIRLDPDINGRYYWRGPEDSGQGLIFQITAF 60

Db 37 LLLLLASLLPSARLASPLPREEEIVFPKLGNSVL-----PGSGAPARLLCRLQAF 88

QY 61 QEDFYLHLTPDAQFLAPAFSTHGLVPLQGLTGGSSDLRRCFYSGDVNAEPDSFAAVSLC 120

Db 89 GETLLEQDSGVQVQGLTVQVQAPL-LGGAP--GTLYTGTINGDPESVSLHWD 145

QY 121 GG-LRCAGFYGAEYVISPLPNASAPAAORNQGAHLQRRGVPGGSDPSTRSGVAG 179

Db 146 GGALLGVLOYRGAELHQLPLEGGTPNSA--GGPGAHILRRK---SPASGGQPMCNV--- 196

QY 180 WNPAILRALDPKPRRAGFSGRSRRSRRAKRFYSIPRYVETLVVADESVMKFGADLE 239

Db 197 -----KAPLGSPSPRPR--RAKRFASLSRFVETLVVADDKMAAFHAGLAK 239

QY 240 HYLLTLATAARLYRHPSILNPNIVVVKVLLLRDRDSGPKVTGNALTLRNFCWAKQK 299

Db 240 RYLLTVMAAAAKAFKHPISIRNPVSLVTVTLVILSGEGPQVGPSSAAQTLRSFCWQRL 299

QY 300 NKVSKHPYWDTAIFLTRQDLCSGATCDTLGMADVTGMDPKRSCSVIEDDGLPSAFTT 359

Db 300 NTPESGDPHFDTAIFLTRQDLCSGATCDTLGMADVTGMDPARSCAIVEDDGLQSAFTA 359

QY 360 AHELGHVFNHNDNVKVCVEFGKL-RANHMSPTLIQIDRANPWSACSAIIITDLDG 418

||||| 360 AHELGHVFNMLHNSKPCISLNGPLSTRHVMAVMAHVDPEPWSPCSAFRTDFLNG 419
QY 419 HGDCLLDQSPKPSISLPEDLPASYSYLSQOCALAFVGSKPCPYM-QYCTKLWCTKAKGO 477
Db 420 YGHCLLDKPEAPLHLPVTFPGDYDADQCLTFGDSRHCPLPPPCAALWCSGHLNGH 479
QY 478 MVCOTRHFHWADGTCGEGKCLKLGACVERHNLNKHVRD--GSWAKWDVPYGPCSRTCGGG 535
Db 480 AMCQTKHSPWADGTPCGPAQACMGRCGLHMDQLQDNIPQAGWGPMGDCSRTCGG 539
QY 536 VOLARRQCTNPTPANGKYCEGVVRYKRSNLEPCPSSASGKSFREOCEAFNGYHNSTN 595
Db 540 VQFSRSDCTRPVPRNGKYCEGRTRFRSCNTEDCP-TGSALTFRBOCAA---YHNRTD 595
QY 596 RLTL--AVAWPKYSGVSPDRCKKLCICRANGTGYFYVLAPKVVDGTLCSPDSTSVCVQG 652
Db 596 LKSPFGPMQWVRYTGVAPQDOCKLTQCARALGYVYVLEPRVVDGTPCSPDSSSVCVQG 655
QY 653 KCICAGCDGNLGSKKRFDKCGVCGDNKSKKVTG 687
Db 656 RCIHAGCDRIIGSKKKFKDKMVCWGGDGGSGCKSQSG 690

RESULT 14

US-10-176-758-352
; Sequence 352, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176, 758
; CURRENT FILING DATE: 2002-06-21
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-352

Query Match 45.1%; Score 1688.5; DB 9; Length 837;
Best Local Similarity 47.2%; Pred. No. 1.6e-120;
Matches 328; Conservative 103; Mismatches 215; Indels 49; Gaps 14;

QY 1 MLLGLITLAFAGTAGGFEPEVVPVIRLDPDINGRYYWRGPEDSGQGLIFQITAF 60
Db 37 LLLLLLASLPSARLASPLPREEEIVPEKLSVLT-----PGSGAPARLLCRLQAF 88
QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRCFYSGDYNAPDPSFAVSLC 120
Db 89 GETLLLEQDSGVQVEGLTVQYLGAPE-LLGGAEP--GTyltGTINGDPESVASLHWD 145
QY 121 GG-LRGAFGYGAEVTSPLPNASAPAAORNSQGAHLQRRGVPGSPGDPSTSRGCVASG 179
Db 146 GGALLGVQYGAELHQLQLEGGTPNSA--CGPAHLIRK-----SPASGGPMCNV--- 196
QY 180 WNPAILRALDPKPRAGFGESRRSRAKRFVSPRYVETILVVADESMKPHGADLE 239
Db 197 -----KAPLGSPSPRPR--RAKREASLSRFVETLWADDKMAAPHGAGLK 239

QY 240 HYLLTLATARLYRHRPSILNIPINIVVVKVLLLRDRSGPKVTGNALTLRNECAQKKL 299
Db 240 RYLLTVMAAAAKAFKHPISIRNPSVTVTRVLVILSGEGEPQVSPAAQTILRSCAMORGL 299
QY 300 NKVSKHPEYWDTAILFTRODLGATTCDTLGMADVGTMDCKPRKSCSVIEDDGLPSAFTT 359
Db 300 NTPEDSGPHDFDTAILFTRODLGATTCDTLGMADVGTMDCKPRKSCSVIEDDGLPSAFTT 359
QY 360 AHELGHVFNMLHNSKPCISLNGPLSTRHVMAVMAHVDPEPWSPCSAFRTDFLNG 419
Db 360 AHELGHVFNMLHNSKPCISLNGPLSTRHVMAVMAHVDPEPWSPCSAFRTDFLNG 419
QY 419 HGDCLLDQSPKPSISLPEDLPASYSYLSQOCALAFVGSKPCPYM-QYCTKLWCTKAKGO 477
Db 420 YGHCLLDKPEAPLHLPVTFPGDYDADQCLTFGDSRHCPLPPPCAALWCSGHLNGH 479
QY 478 MVCOTRHFHWADGTCGEGKCLKLGACVERHNLNKHVRD--GSWAKWDVPYGPCSRTCGGG 535
Db 480 AMCQTKHSPWADGTPCGPAQACMGRCGLHMDQLQDNIPQAGWGPMGDCSRTCGG 539
QY 536 VOLARRQCTNPTPANGKYCEGVVRYKRSNLEPCPSSASGKSFREOCEAFNGYHNSTN 595
Db 540 VQFSRSDCTRPVPRNGKYCEGRTRFRSCNTEDCP-TGSALTFRBOCAA---YHNRTD 595
QY 596 RLTL--AVAWPKYSGVSPDRCKKLCICRANGTGYFYVLAPKVVDGTLCSPDSTSVCVQG 652
Db 596 LKSPFGPMQWVRYTGVAPQDOCKLTQCARALGYVYVLEPRVVDGTPCSPDSSSVCVQG 655
QY 653 KCICAGCDGNLGSKKRFDKCGVCGDNKSKKVTG 687
Db 656 RCIHAGCDRIIGSKKKFKDKMVCWGGDGGSGCKSQSG 690

RESULT 15

US-10-175-737-352
; Sequence 352, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175, 737
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-352

Query Match 45.1%; Score 1688.5; DB 9; Length 837;
Best Local Similarity 47.2%; Pred. No. 1.6e-120;
Matches 328; Conservative 103; Mismatches 215; Indels 49; Gaps 14;

QY 1 MLLGLITLAFAGTAGGFEPEVVPVIRLDPDINGRYYWRGPEDSGQGLIFQITAF 60
Db 37 LLLLLLASLPSARLASPLPREEEIVPEKLSVLT-----PGSGAPARLLCRLQAF 88
QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRCFYSGDYNAPDPSFAVSLC 120
Db 89 GETLLLEQDSGVQVEGLTVQYLGAPE-LLGGAEP--GTyltGTINGDPESVASLHWD 145

QY 121 GG-LRGAFYRGAEYVISPLPNASAPAAQNSOGAHLLORRGVPGGSDPTSRCGVASG 179
Db 146 GGALLGVQYRGAEHLQPLEGGTPNSA--GGFGAHLRRK-----SPASGGQPMCNCV--- 196
QY 180 WNPAILRALDPYKPRRAGESRRRRSGRAKRFVSIPIRYVETLVVVADESMVKFHHGADLE 239
Db 197 -----KAPLGSPSPRPR--RAKRFASLSRFVETLVVVADESMVKFHHGAGLK 239
QY 240 HYLLTLATAARLYRHPSILNINIVVKKLLRRDSDGPKVTGNAALTLRNFCAWQKKL 299
Db 240 RYLLTYMAAAAKAFKHPISRNVPVSLVTRVLGSGEGEPQVGPSSAAQTLRSFCAWQRL 299
QY 300 NKVSDRHPYWDTAILFTQDLGATTCDTLGMADVTMCDPKRSCSVIEDDGLPSAFTT 359
Db 300 NTPEDSGPHDTAILFTQDLGCGVSTCDTLGMADVTMCDPARCAIVEDDGLQSAFTA 359
QY 360 AHELGHVFNHNDNVKVEEVEFKL-RANHMSPTLIQIDRANPWSACSAAIITDFLDG 418
Db 360 AHELGHVFNHNDNSKPCISLNGPLTSRHMVAPVMAHYDPEPWSPCSAREITDFLDNG 419
QY 419 HGDCILLDQSPKPSLPEDLPGASYTLSQOCELAFGVGSKPCPYM-OYCTKLMCTGKAGQ 477
Db 420 YGCHLLDKPEAPLHLPVTPFGKDYDADROQLTFGPDNRHCPQLPPPCAAALWCSGHLNGH 479
QY 478 MVCOTRHFWDGTSGGEGKCLKGACVERHNLNKHRYD--GSAKWDPYGPSCSRTCXGG 535
Db 480 AMCQTHSPWADGTPCGPAQACWGRCLHMDQLQDFNIPQAGGWPWGPWGDGCSRTCXGG 539
QY 536 VOLARQCTNPTPANGKYCEGVYKYSRSCNLEPCPSSASGKSFREEQCEAFNGYNHSTN 595
Db 540 VQFSSRDCTRPVPRNGKYCEGRRTRFRSCNTEDCP-TGSALTFREEQCAA---YNHRTD 595
QY 596 RLTL---AVAWPKYSGVSPRCKLICRANGTGYFYVLAPKVVDGTLCSPDSTSVQVQ 652
Db 596 LFKSFPQMDWVPRYTGVAPOQCKLTQCARALGYYYVLEPRVVDGTPCSPDSSSVQVQ 655
QY 653 KCIKAGCDGNLGSKKRFDKCGVCGDNKSKKVTG 687
Db 656 RCIHAGCDRIIGSKKFKDKMVCVCGDGCCKSQSG 690

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 15.8904 Seconds
(without alignments)
4156.253 Million cell updates/sec

Title: US-10-009-332-1_COPY_1_687

Perfect score: 3744

Sequence: 1 MLLLGILTLAFAGTAGGFE.....RFDKCGVCGDNKSKKVTG 687

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1947.5	52.0	951	2 T00017	gene ADAMTS-1 prot
2	1687.5	45.1	837	2 T00355	hypothetical prote
3	1223	32.7	2165	2 T21371	hypothetical prote
4	1084	29.0	550	2 T47158	hypothetical prote
5	980.5	26.2	1205	2 T18517	procollagen N-endo
6	479	12.8	1444	2 T18856	angiogenesis inhib
7	426.5	11.4	860	2 T16892	hypothetical prote
8	406.5	10.9	1558	2 C89114	protein C37C3.6a l
9	406.5	10.9	2167	2 T34395	hypothetical prote
10	362	9.7	957	2 T15976	hypothetical prote
11	344.5	9.2	571	2 S24789	jatrahagin C precu
12	337	9.0	609	2 S55270	catrocollastatin p
13	329	8.8	903	2 S60257	metrin alpha - mo
14	323.5	8.6	617	2 S48160	metalloproteinase
15	313	8.4	549	2 S48169	metalloproteinase
16	313	8.4	826	2 A60385	monocyte surface a
17	309.5	8.3	951	2 T00260	hypothetical prote
18	307.5	8.2	789	2 S28259	androgen-regulated
19	307	8.2	411	1 HYSNFA	fibrolase (EC 3.4.
20	304	8.1	407	2 S66260	metalloproteinase
21	300.5	8.0	616	2 A55796	ecarin precursor -
22	297	7.9	480	1 A30065	trigamin precursor
23	292	7.8	478	2 JC4880	fibrinolytic metal
24	290.5	7.8	481	2 JC4342	fibrinolytic prote
25	289	7.7	414	2 S41609	atrolysin C (EC 3.
26	286.5	7.7	610	2 JC7530	vascular apoptosis
27	284	7.6	414	1 HYSAC	atrolysin C (EC 3.
28	280	7.5	481	2 S43125	trimucin precursor
29	277.5	7.4	814	2 G02390	disintegrin-like m

30	273.5	7.3	478	2 A43296	atrolysin E (EC 3.
31	265	7.1	414	2 S41608	atrolysin B (EC 3.
32	258.5	6.9	670	2 I65967	disintegrin-like m
33	257	6.9	776	2 S28258	androgen-regulated
34	251	6.7	1042	2 T26644	hypothetical prote
35	248.5	6.6	1059	2 T22545	hypothetical prote
36	248.5	6.6	1170	1 TSHUP1	thrombospondin 1 p
37	245.5	6.6	1170	2 A40558	thrombospondin 1 p
38	244.5	6.5	419	2 A59414	metalloproteinase
39	244	6.5	952	2 T18900	disintegrin and me
40	243	6.5	478	2 JQ1301	hemorrhagic protei
41	234.5	6.3	508	2 T22836	hypothetical prote
42	233.5	6.2	419	2 S41607	atrolysin A (EC 3.
43	232	6.2	823	2 S18968	cyritestin precurs
44	228	6.1	735	2 G02937	fertilin beta - cr
45	220	5.9	416	2 A37877	hemorrhagic protei

ALIGNMENTS

RESULT 1

T00017

gene ADAMTS-1 protein - mouse

C:Species: Mus musculus (house mouse)

C>Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000

C:Accession: T00017

R:Kuno, K.; Lizasa, H.; Ohno, S.; Matsushima, K.

Genomics 46, 466-471, 1997

A:Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 g

A:Reference number: Z14055; MUID:98110583; PMID:9441751

A:Accession: T00017

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-951 <KUN>

A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BAA24501.1; PID:g2809057

A:Superfamily: thrombospondin type 1 repeat homology

C:Experimental source: strain 129SVJ

C:Genetics:

A:Gene: ADAMTS-1

A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 602/1; 660/3; 719/2

F:542-598/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match

Best Local Similarity 52.0%; Score 1947.5; DB 2; Length 951;

Matches 381; Conservative 95; Mismatches 176; Indels 77; Gaps 16;

QY 1 MLLLGILTLAFAGTAGG--FEPEREVVPVIRLDPDINGRRYWRGP-EDSGDQGLIFOI 57

Db 20 LLLASITMLLCARGAHRPTDEELVLP-SLE-----RAPHGDSITTRL--RL 66

QY 58 TAFQEDFYHLTPDAQFLAPAFSTHGLVPLQGLTGG-----SDLRRCFYSGDVNA 109

Db 67 DAFQOOLHLKLPDPSGLFAPGTLQTV-----GRSPGEAQLHLDPTGLAHCFYSGTVNG 121

QY 110 EPDFAAAYSLCGLRGAFYGAAYVISPFPNAS-----APAAQRNSQGA----HLLORRG 161

Db 122 DPGSAALSLSCEVGAFYLOGEEFFIQAPGVATERLAPVPEESSARFQHLRRR- 180

QY 162 VPGSGSGDPTSCGVASGNPAILRALDPYKPRFAGFGESESRRR----- 206

Db 181 ----RSGSGAKCGYMD-----DETLPDTSRPSQNTNRQWVPVDRPTQDACKP 226

QY 207 ---SGRAKRFVSIPIRYETLWVADESMVKFHGADLEHYLLTLTAARLYRHSILNPI 262

Db 227 SGPGSIRKKRVSSPRYYKFTMLVADQSMADFHGSLKHYLLTLFSVAARFYKHSIRNSI 286

QY 263 NIVVYKVLRLDRDGPVKVTGNAALTNRNCAWKKNKSDKHPEYWDITAILTRDLC 322

Db 287 SILVVYKILVIYEQKGPVTSNAALTNRNCAWKKNKSDKHPEYWDITAILTRDLC 346

QY 323 GATTCDDITGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFG 382

Db 323 GATTCDDITGMADVGTMCDDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFG 382

Db 347 GSHTCDTLGMADVGTVCDFSRSCSVIEDDGLQAFAAHTAHELGHVFNMPHDDAKHCASLNG 406
 QY 383 KLRANHMWSPPTLQIDRANPWSACSAIIITDFLDGSHGDCILDDQPSKPTSLPDLPGASY 442
 Db 407 VTGDSHLMASLSSLDHSPQSPSCSYWYTSFDLNGHGCLMDKPKONPKLPSDLPGTLY 466
 QY 443 TLSQOCLEAFAGSKPCP-YMOTYCTKLWCTGKAGQWQVQTRHFPWADGTCGEGKLCUK 501
 Db 467 DANROCOFTFEGESKHCPOAASCTTTLWCTGTSGGLLVQTKHFPWADGTCGEGKWCVS 526
 QY 502 GACVERHNLNKH---RVDSNAKWDYGPSCSTCGGGVQVQLARRQCTNPTPANGKYCEGV 558
 Db 527 GKCVNKTDM-KHEATPVHGSWGPWGDCSRCTCGGGVQVYTMRECDNPPKNGKYCEGK 585
 QY 559 RVKYSRNLPCPSSASGSKFREEOCEAFNGYHNSNRLTLAVAWPKYSGVSPDRCKL 618
 Db 586 RVKYSRNLPCPSSASGSKFREEOCEAFNGYHNSNRLTLAVAWPKYSGVSPDRCKL 644
 QY 619 ICRANGTGYFVLAPKVVDTGLSPDSTSVQVQKCIKAGCDNGLSKKREDCGVCVGD 678
 Db 645 TCEAKGIGYFVLQPKVVDGTPCSPDSTSVQVQCVKAGCDRIIDSKKFKDKGVCVGN 704
 QY 679 NKSCCKVTG 687
 Db 705 GSTCKMKG 713

RESULT 2
 T00355
 hypothetical protein KIAA0688 - human
 C:Species: Homo sapiens (man)
 C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
 C:Accession: T00355
 R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
 DNA Res. 5, 169-176, 1998
 A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete
 A:Reference number: Z14142; MUID:98403880; PMID:9734811
 A:Accession: T00355
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-837 <ISH>
 A:Cross-references: EMBL:AB014588; NID:g3327189; PIDN:BA31663.1; PID:g3327190
 A:Experimental source: Brain
 C:Genetics:
 A:Gene: KIAA0688
 A:Superfamily: thrombospondin type 1 repeat homology
 F:519-575/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 45.1%; Score 1687.5; DB 2; Length 837;
 Best Local Similarity 47.2%; Pred. No. 5.5e-112; Mismatches 215; Indels 49; Gaps 14;
 Matches 328; Conservative 103;

QY 1 MLLGILTLAFAGTAGFEPEREVVPIRLDPDINGRRYWRGPDSDGQGLIFQITAF 60
 Db 37 LLLLLLASLPSARLASPLPREIEIVPEKLSVLF-----PGSGTTPARLLCRLOAF 88
 QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVLPLQGLTGSSDLRCFCYSGDYNAPDFAAAYSLC 120
 Db 89 GETLLELEQDSGVQVEGLTVQYLGQAPL-LLGGAEP--GTYLGTINGDPESVASLHWD 145
 QY 121 GG-LRGAFGYGAEVVISPLPNASAPAAORNSQGAHLORRGVPGPGSDPTSRGCVASG 179
 Db 146 GGALLGVLOYGAELHQLPLEGGTFNSA--GGPGAHILRRK-----SPASGGPMCNV--- 196
 QY 180 WNPAILRALDYPKPRAGGESRRRRSGRAKRFVSPRYVETVLVVADESMVKFHGADLE 239
 Db 197 -----KAPLGSPSPRPR--RAKREASLSRFVETLVVADDDKMAAFHAGLAK 239
 QY 240 HYLLTLLAARLYRHPHSILNPINIVVYKVLRLDRDGGPKVTGNAALTNRNFCWOKKL 299
 Db 240 RYLLTVMAAAAKAFKHPHSIRNPVSLVRLVILSGEGEGPQVGFSAQTLRSFCWOKGL 299
 QY 300 NKVSDKHPEYWDITLFTFQDLGATTCDTLGMADVGTVMCDPKRSCSVIEDDGLPSAFTT 359

Db 300 NTPEDSDPDHEDTALLTRQDLQGVSTCDILGMADVGTVCDFSRSCSVIEDDGLQSAFTA 359
 QY 360 AHELGHVFNMPHNDVNVKVEEYFGKL-RANHMWSPPTLQIDRANPWSACSAIIITDFLDG 418
 Db 360 AHELGHVFNMLHDNSKPCISLNGPLSTSRHVMAPVMAHVDPEEPWPCSFARFIDFLDNG 419
 QY 419 HGDCLLDQSPKPSLPELPGASTYLSQOCLEAFAGSKPCPYM-QYCTKLWCTGKAKGQ 477
 Db 420 YGHCLLDPEAPLHLPVTFPGKYDADQCOLTFGDSRHCPLPPPCAALWMSGHLNGH 479
 QY 478 MVQOTRHPADGTCGEGKLCAGACVERHNLNKHRYD--GSNAKWDYGPSCSTCGGG 535
 Db 480 AMCQTKHSPWADGTPCGPAQACMGRCRLHMDQLQDFNIPQAGGPGWPGWDCSRTCGG 539
 QY 536 VOLARRQCTNPTPANGKYCEGVRYKYSRNLPCPSSASGSKFREEOCEAFNGYHNSN 595
 Db 540 VQFSSRDCTRPVPRNGKYCEGRRTRFRSCNTEDCP-TGSALTFRFEEQCAA---YHRTD 595
 QY 596 RLTL---AVAWPKYSGVSPDRCKKLCICRANGTGYFYVLAPKVVDTGLSPDSTSV 652
 Db 596 LFKSPFGPMWVPRYTGVAPODQCKLTQARALGYVYVLEPRVVDGTPCSPDSSSV 655
 QY 653 KCIRKAGCDNGLSKKREDCGVCVGDNKSCKKVTG 687
 Db 656 RCIRHAGCDRIIGSKKKFKDKCVCVGDGSGCKSKQG 690

RESULT 3
 T21371
 hypothetical protein F25H8.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T21371; T24896
 R:Gajadaty, S.
 submitted to the EMBL Data Library, February 1996
 A:Reference number: Z19413
 A:Accession: T21371
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2165 <W1>
 A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
 A:Experimental source: clone F25H8
 R:Gajadaty, S.
 submitted to the EMBL Data Library, February 1996
 A:Reference number: Z19949
 A:Accession: T24896
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2165 <W1>
 A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
 A:Experimental source: clone T13H10
 C:Genetics:
 A:Gene: CESP:F25H8.3
 A:Map position: 4
 A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/2; 634/2; 744/1;

Query Match 32.7%; Score 1223; DB 2; Length 2165;
 Best Local Similarity 40.5%; Pred. No. 1.7e-78;
 Matches 257; Conservative 87; Mismatches 214; Indels 76; Gaps 16;

QY 101 CFYSGDVNAEPDFAAVALSLC---GGLRGAFGYGAEVVISPLPNASA---PAAORNSQA 154
 Db 169 CIYRAHVKG-VHQHSIVNLCDSGDLGYLALPSGIHTVEITSGNGTEHDGASRHRQ-- 225
 QY 155 HLQORGVPGGSDPTSRGCV-----ASGNPAILRALDYPKPRAGGESRRSR 204
 Db 226 HLVRKFDPMHFKSFDHLNSTSVNETETTVATWQDQEDVIER-----KARSR 272
 QY 205 RRSGRKRRFYSIPRYVETVLVVADESMVKFHGADLEHYLLTLLAARLYRHPHSILNP 264
 Db 273 RAAN-----SWDHYVEVLVVAATKMYEYHGRSLDEYVLTFTVTSIYRHQSIRASIN 326

—
—
—

—
•
•
—

•
•
—

•
•
—

—
•
•
•
•
•

Db 382 ---KSDTLFTGTCGSKWQGLRCVPTWCTNEIQPTVQHVAPVVTLPSPRIDGWSG 438
 QY 421 DCLDQ-----PSKIPSLPDLPGAS-----YTLSSQCELA----- 451
 Db 439 ATICSQCTNGILGVLGATARTCSAPYPANGSGDCVSTRAVLCRSQCGRASKSVDE 498
 QY 452 -----FVGSKPCPYMYQYCKLMCTGKAKGQVQCTRHFPW-----A 488
 Db 499 YISDKMEQKRLKNDRELTGKGLNRFQPCACKVFCVQDQ---QHYGSRNYRFFGDNLP 555
 QY 489 DGTSCGEGKCLLKGACVERINLKHVD-----GSAWAKDYPGP 527
 Db 556 DGTSCGVDYCLDCEALCNMNLISRDQSCPTDTCPTDSSSVYRGQWGTSLWTS 615
 QY 528 CSRTCGGQVLARRQCTNTPANGGKYCEGVVRKYRSCNLEPCP----- 571
 Db 616 CTATCGGGRKNRACS---ITG---QCEGNEDETEVCSESCPSVLRVGNESWTWEN 669
 QY 572 ---SSASGKSFRQEC-----EAFN---CYNHSTNRLTLAVAWVPKYSGVSPRDCKLI 619
 Db 670 HCSVSCGRGQARYRKCLSPHRTLAFDCGENKVTNELRTFFKARSYIMCSVR---CNKI 727
 QY 620 CR-----ANGTGYFVYVAPKVVVDGTLCSPDSTSVCVGKCIKACDGN 662
 Db 728 KRNTISEKNIEVRSCDNGPCNAIGVWGTGWSGTCSTSCGPG---TLVRQTCNREPCDGS 785
 QY 663 LGSKKRFDKGV---CGDN-----KSKCKVTG 687
 Db 786 AHERR---SCNVATCQNDGIWLSLWNEWSDCSRVCG 817

RESULT 11

S24789
 Jaraahagin C precursor - Jaraahaca (fragment)
 N:Alternate names: single chain botrocetin
 C:Species: Bothrops jaraahaca (Jaraahaca)
 C>Date: 20-Feb-1995 #sequence_revision 29-Aug-1997 #text_change 09-Jun-2000
 C:Accession: S24789; J2245; A4463; A37958; J22373
 R:Paine, M.J.I.
 submitted to the EMBL data library, August 1992
 A:Reference number: S24789
 A:Accession: S24789
 A:Molecule type: mRNA
 A:Residues: 1-571 <PAI>
 A:CROSS-references: EMBL:X68251; NID:g62467; PID:g62468
 R:Usami, Y.; Fujimura, Y.; Miura, S.; Shima, E.; Yoshida, E.; Hirano, K.;
 Biochem. Biophys. Res. Commun. 201, 331-339, 1994
 A>Title: A 28 kDa-protein with disintegrin-like structure (Jaraahagin-C) purified from B
 A:Reference number: J2245; MUID:94256999; PMID:8198592
 A:Accession: J2245
 A:Molecule type: protein
 A:Residues: 360-571 <USA>
 A:Experimental source: venom
 R:Paine, M.J.; Desmond, H.P.; Theakston, R.D.; Crampton, J.M.
 J. Biol. Chem. 267, 22869-22876, 1992
 A>Title: Purification, cloning, and molecular characterization of a high molecular weight
 ily.
 A:Reference number: A44463; MUID:93054601; PMID:1385408
 A:Accession: A44463
 A:Molecule type: mRNA
 A:Residues: 1-23, 'Q', '94-131, 'G', '133-169, 'Q', '171-571 <PA2>
 A:CROSS-references: GB:X68251
 A:Experimental source: venom gland
 A>Note: sequence inconsistent with the nucleotide translation
 R:Fujimura, Y.; Titani, K.; Usami, Y.; Suzuki, M.; Oyama, R.; Matsui, T.; Fukui, H.; Sug
 Biochemistry 30, 1957-1964, 1991
 A>Title: Isolation and chemical characterization of two structurally and functionally di
 A:Reference number: A37958; MUID:91129280; PMID:1993206
 A:Accession: A37958
 A:Molecule type: protein
 A:Residues: 360-372, 'E', '374-378, 'X', '380-384 <FUW>

A:Note: 361-Val was also found
 C:Comment: Inhibits collagen- and ADP-induced platelet aggregation.
 C:Superfamily: mouse meltrin alpha; disintegrin homology
 C:Keywords: hydrolase; metalloproteinase; venom; zinc
 F:360-571/Product: Jaraahagin C #status experimental <MAT>
 F:362-444/Domains: disintegrin homology <DIS>
 F:295,299,305/Binding site: zinc (His) #status predicted
 F:296/Active site: Glu #status predicted

Query Match 9.2%; Score 344.5; DB 2; Length 571;
 Best Local Similarity 22.7%; Pred. No. 9.3e-17;
 Matches 160; Conservative 88; Mismatches 256; Indels 201; Gaps 35;

QY 12 AGRTAGGPEPEVVPVPIRLDPDINGRIRYWRGPDSDGQLIFQITAFQEDF-YLHLTP 70
 Db 1 ATRPKGAVQPKYEDA--MQYEFKNGEPVWLHLEKNKG-----LFSKDYSEIHYP 49
 QY 71 DAQFLAPAFSTHGLVPLQGLTGGSSDLRCFYSDVNAEPDSPAFAVSLCGLGCAFGYR 130
 Db 50 DGREITTPPVED-----HCYHGRINDADSTASISACNLGKYFKLQ 93
 QY 131 GAQYVISP--LPNASAPAAQNSQGAHLLQRRGVPDGPSPDTSRCGVASGN----- 181
 Db 94 RETYFIEPLKLPDSEHA-----VFKYENVE--KEDEAPKMGVYQNKWSVEPIKK 142
 QY 182 -----PAILRALDPYKPRRAGFGESRRSRRAKRVSPVRYVETLVVADESVMKPHG 235
 Db 143 ASOLAFATQQRYPDK-----YIEFFVVDQGTVTTKNN 176
 QY 236 ADLEHYLLTLATAARLYRHPSILNPI--NIYVVKVLL--LRDRSGPKVT--GNAALTLR 290
 Db 177 GOLD-----KIKARMTYELANIVNEIFRYLYMHVALVGLNWSNGDKITVDPDYTLN 229
 QY 291 NFCAQKKLNVSKHPEYWDITAILFTRODLGCAFTCTDGLMADYGTMCDPKRSQSVIED 350
 Db 230 SPAEWRKTDLLTRKKH---DNAQLLTAFDENG---PTIGVAYIGSMCHPKRSVGIQVD 281
 QY 351 ----DGLPFAFTAHGLHVFVNMHNDVNVKVEVFGKLRANH---MMSPTLQIDRANPWS 404
 Db 282 YSPINLVAVIMAHENMGNLGIHDT-----GSCSGDYPCIMGPT--ISNEPSKFFS 332
 QY 405 ACSAAIITDLDSDGHGDCLLDOP-----SKPISLPDLPGASVYTLSSQCELAFLGVS 458
 Db 333 NCSYIQCWDFIMNHNPCEIINEPLGTDIISPPVCGNEL-----EVGECD----- 378
 QY 459 CPYMOYCTKLWCTGKAGQVQCTRHFPWADTSCGEGKCLKGACVERHNLNKHVD-- 516
 Db 379 CQTPENCNECCDA-----ATCKLK-----SGSQCGHGD-----CCBQCKFSKGTBCR 422
 QY 517 GSAWAKDYPGCSRTCGGQVLARRQCTNTPANGGKYCEGVVRKYRSCNLEPCP----- 571
 Db 423 ASMSECDFAEHCCTGSS-----ECPADVPHKNGQPCLD---NYGYCYNGNCPIMVHQ 471
 QY 572 -----SSASGKSFRQECFAENGYNHSTNRLTLAVAWVPKYSGVSPRD-KC-KLIC 620
 Db 472 CYALFGADVYEEDSCFDKDNQKNGYVYCRKNGKKIIFCA-----PEDVKCGRLYC 522
 QY 621 RANGTG-----YFYVLAPK-----VVDGTLCSPDSTSVCVGKCI 655
 Db 523 KNSPQNNPCKMFYSNDDEHKHGMVLPCTKCA--DGKVCNSNGHCV 565

RESULT 12

S55270
 catrocollastatin precursor - western diamondback rattlesnake
 C:Species: Crotalus atrox (western diamondback rattlesnake)
 C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
 C:Accession: S55270
 R:Zhou, Q.; Smith, J.B.; Grossman, M.H.
 Biochem. J. 307, 411-417, 1995
 A>Title: Molecular cloning and expression of catrocollastatin, a snake-venom protein
 A:Reference number: S55264; MUID:95251603; PMID:7733877
 A:Accession: S55270

A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-609 <2HO>
A;Cross-references: GB:U21003; NID:g710353; PIDN:AAC59672.1; PID:g710354
C;Superfamily: mouse meltrin alpha; disintegrin homology

```
Query Match      9.0%; Score 337; DB 2; Length 609;
Best Local Similarity 22.8%; Pred. No. 3.4e-16;
Matches 150; Conservative 90; Mismatches 259; Indels 160; Gaps 30;

QY 50 DOGLIFQITAFQEDFYHLTPDAQFLAPAFSTEHLGVPLOGLTGGSSDLRRCFYSGDNA 109
Db 52 EDAMQVELKYNQEPVVLHLKNGKLFSDYSETHYSPDGREITTYPLVEDHCVYHGRIEN 111
QY 110 EPDSFAVSLCGLRGAFYRGAEYVISP--LPNASAPAAQNSQGAHLQRRGVPGPS 167
Db 112 DADSTASISACNLKGKHFQLOGBMYLIEPLKLPDSEAHAVKYENVE-----KE 160
QY 168 GPTSRCGVASGNPAILRALDPYKPRRAGFGESRRRSRGRKRFVSIIP-----RY 219
Db 161 DEALKMGVTONW-----ESYEPIK-----KASQLVVTAEHQKYNPFP 199
QY 220 VETLVVADESVMKFGHADLEHY---LLTLATAARLYRHPSTILNINIVVVKLLLRDRD 276
Db 200 VELFLVVDKAMVTKNNGDLDKIKRMYEIVNTVNEYIRYMY---IHVALVGLIWSNED 255
QY 277 SGPKVIT--GNAALTIRNCAWOKKLNKYSKHPYWDTAILFTRODLGCATTCDTLGMAD 334
Db 256 ---KIIVKPEAGYTLNAGFGEWKTDLLTRKH---DNAQLLTAIDL-----DRVIGLAY 303
QY 335 VGMCDPKKSCSVIED---DGLPSAFTTAHELGHVFNMPHDN-VKVCBEVFGKLRANHMM 390
Db 304 VGSWCHPKRSTGLIQDYSEINLVAVIMAHENGHNIGINDHSDYSCSGDYACIMRPEISP 363
QY 391 SPITLIQIDRANPSACSAALITDLDSDGDCDLDLP-----SKPISLPDLPGASYTL 444
Db 364 EPSTF-----FNSCKYFECDWDFIMNHNPICILNPLGTDIISPPVCGNELL-----EV 411
QY 445 SOCELAFAFGVSGKPCPMYCYKLTCTGKAKGMVQCOTRHPFWADTSCGEGKCLKGKAC 504
Db 412 GEBGD-----CGTPENCNECDA-----ATCKLK-----SGSQCGHGD-----C 446
QY 505 VERHNLNKHURVD--GSWAKWDYRPGCSTRCTCGGVOLARQCTNPPTPANGKYCEGVRYKY 562
Db 447 CEQCKFSKSGTECRASMSCECPAEHCTGSS-----ECPADVFEHKNQGPCLD---NY 495
QY 563 RSNCLBPCC-----SSASGKSFREOCEAFNGYHNSTNRLTLAVANVPKYS 608
Db 496 GYCYNGNCPIMHQCYDLEFAGADYVEAEDSCFERNKGNFYGYCRKENGKNKIPCA----- 549
QY 609 GVSPRD-KC-KLICRANGTG-----YFYVLAPK-----VVDGTLCSPDSTSVCVQKCI 655
Db 550 ---PEDVKCGRLYCKDQNSPGQNNPKMFYSNEDEHKHMYLPCTKCA--DGKVCNSGHCV 603
```

RESULT 13

S60257
meltrin alpha - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 20-Jun-2000
C;Accession: S60257
R;Yagami-Hiromasa, T.; Sato, T.; Kurisaki, T.; Kamiyo, K.; Nabeshima, Y.I.; Fujisawa-Seki
Nature 377, 652-656, 1995
A;Title: A metalloprotease-disintegrin participating in myoblast fusion.
A;Reference number: S60257; MUID:96026308; PMID:756181
A;Accession: S60257
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-903 <YAG>
A;Cross-references: EMBL:D50411; NID:g1054586; PIDN:BAA08912.1; PID:g1054587
C;Superfamily: mouse meltrin alpha; disintegrin homology
F.421-503/Domain: disintegrin homology <DIS>
F.349/Active site: Glu #status predicted

Query Match 8.8%; Score 329; DB 2; Length 903;
Best Local Similarity 21.5%; Pred. No. 2e-15;
Matches 159; Conservative 74; Mismatches 245; Indels 262; Gaps 27;

```
QY 2 LLLGLTLTAFAGRTA-----GGFEPEREVVVPIRLDPDINGRRYWRGPEDSGDQG 52
Db 15 LLLLAGALLAPRAARGSLMDQRGAYEVARSL--LSKDPGIPGQSI-----PAKDHPDV 68
QY 53 LIFQITAFQEDFYHLTPDAQFLAPAFSTEHLGVPLOQ-----GLTGGSSDLRRCFYSGDV 107
Db 69 LTVQLQLESRDLLSLRNEGLIANGFTETHY---LDQTDVSLTRNHTD--HCYHGHV 123
QY 108 NAEPSFAVSLCGLRGAFYRGAEYVISP--LPNAS-----APA-AQRNSQGAHLQRRG 161
Db 124 QGDAASVYSLTSCDSLRLIMFENKTYSLPEMKNTTDSYKLVPAESMTNIQCL----- 176
QY 162 VFGSPGSDPTSRGCVASGNPAILRALDPYKPRRAGFGESRRRSRGRKRFVSIIPRYVE 221
Db 177 -----CSQHNKSNLTWEDVSP-----GTSQMRARRHK---RETLKMTKYVE 215
QY 222 TLVVADESVMKFGHADLEHYLLTLATAARLYRHPSTILNINIVVVKLLLRDRDSGPKV 281
Db 216 LVIVADNREFQKQKLEKVKORLIEIANHVDKFRPLN-IRIVLVGVEVWINDIK-CSI 273
QY 282 TGNAAITLRNCAWOKKLNKYSKHPYWDTAILFTRODLGCATTCDTLGMADYGTWCDP 341
Db 274 SODPFTRLHEFLDWKRIKLLPRKSH---DNAQLISGYVFOGT-----TIGNAPIMSMCTA 325
QY 342 KRSCSVI---EDDGLPSAFTTAHELGHVFNMPHDN-VKVCBEVFGKLRANHMMSPTLIQI 397
Db 326 EQSGGVNMDHSDSPGAAVTIAHELGHVFNHHTLERGCSRMAAEKGGCIMPSS----- 381
QY 398 DRANP---WSACSAALITDLDSDGDCDLDLP-----SKPISLPDLPGAS----- 427
Db 382 -TGPFPMVFSSCKDLEASLEKGMGMLNPLVEQAQFGKRGKNGYVEEGECCDGE 440
QY 428 -----SKPISLPDLPGAS-----SKPISLPDLPGAS----- 441
Db 441 PEECTNRCCNATTTLKPDVAVCAHQCCEDCQLKPPGTACGSSNSCDLPEFCTGTAPHC 500
QY 442 -----YTLSSQCELAFAFGVSGKPCPMYQV----- 464
Db 501 PANVYLDHGHPCQGVGYNGYICQTHEQOCVTLWGFPAKPAICFCIFERNVSAGDPYGN 560
QY 465 -----CTKLWCTGKAK-----GOMVQCOTRHPFW 487
Db 561 GKDSKSAFAKCELRDAKCGKIQCGQGSARPVIGTNAVSIETNIPQOEGGRILCRGTHVL 620
QY 488 AD-----GTSCGEGKCLKGACVERHNLNKHURVDGWSAKWDYRPGCSTRCTCGGV 536
Db 621 GDDMPDPLVLAGTKCAEGKICLNRCQNISVFGVHK-----CAMQCHGRGV 667
QY 537 QLARRQCTNPTPPANGKYCE 556
Db 668 CNNRKNC-----HCE 677
```

RESULT 14

S48160
metalloprotease (EC 3.4.24.-) H-I precursor - carpet viper
N;Contains: disintegrin
C;Species: Echis pyramidum leakeyi
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jul-2000
C;Accession: S48160
R;Palme, M.J.I.; Moura-Da-Silva, A.M.; Theakston, R.D.G.; Crampton, J.M.
Eur. J. Biochem. 224, 483-488, 1994
A;Title: Cloning of metalloprotease genes in the carpet viper (Echis pyramidum leakeyi)
A;Reference number: S48160; MUID:95010025; PMID:7925363
A;Accession: S48160
A;Molecule type: mRNA
A;Residues: 1-617 <PAI>
A;Cross-references: GB:X78970; NID:g763092; PIDN:CAA55565.1; PID:g763093

C:Superfamily: mouse meltrin alpha; disintegrin homology
C:Keywords: hydrolase; metalloproteinase; venom
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-192/Domain: propeptide #status predicted <PRO>
F:193-617/Product: metalloproteinase H-II #status predicted <MAT>
F:403-485/Domain: disintegrin homology <DIS>
F:337/Active site: Glu #status predicted

```
Query Match      8.6%; Score 323.5; DB 2; Length 617;
Best Local Similarity 23.2%; Pred. No. 3.1e-15;
Matches 150; Conservative 74; Mismatches 267; Indels 155; Gaps 27;

Qy 45 PEDSGQGLIFQITAFQEDFYHLTPDAQFLAPAFSTEHLGVLPGLGGSDLRRCRYS 104
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 50 PEQYEDTMRVEKNGVEVLHLEKNGLFSEDIYSETHYSPDGEITNPPVEDHCYH 109
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 105 GOVNAEPDSFAVSLCGGLRGAFGAGYEVISPL--PNASAPAAQRNSQGAHLRQGV 162
    | : : : : || : : : || : : : || : : : || : : : || : : :
Db 110 GRIQNDADSTASACNLGKGYFMLRGETYLIEPLKIPDSEAHAYKYENVE----- 161
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 163 PGSPSGDPTSRGVA--SGWNPAILRALDPYKPRRAGFSGSRRRSRGRAKREVSIPRYVE 221
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 162 ---KEDEAPKCGVTGNES-----DELKASQIVATSEQR-----SYKKIE 203
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 222 TLVVADESMVKFGADLEHYLLTLATAARLYRHPISILPINIVV-----VKVLLR 273
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 204 FVVVADYIMRYKNND-----STAVRRRIEIVNLMVMYIVFNHVALTHIEIWSTR 256
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 274 DRDSGPKYTGNAALTLRNFCAQKLNKYSKHPEDWDFAILFTRODLGGATTCDTLGN 333
    | : : : : || : : : || : : : || : : : || : : : || : : :
Db 257 DQIT---VOSAADVTLDFGDRAKNLLTRKKH---DNAQLFTGINLNG-----OTL 305
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 334 DVGTMCDPKRSQVIED---DGLPSAFTTAHELGHVFNPHDNVNV--CEVEFGKLRANHM 389
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 306 RMKGMSPNSVGVIOQYCKNVLVVAITWAHELGNLGHNDHNGNCNCPDTSICINSAVAG 365
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 390 MSTLIQIDRANPWSACSAIITDFLDGSHGDCLLDQPSKPTSLPEDLPASVY--LSQOC 448
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 366 PEPVF-----SFSNCRNDYRSFRNSDQSKIDNKLKLTQIVSPVCGNVEVEGEC 418
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 449 ELAFVGSKPCPYMOCYCTKLWCTGKAGQWVCQTRHFPWADGTSCGEGKCLKGACVERH 508
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 419 D-----CGSRVTCRNPCCNA-----TTCKL-----TPGSCADGEC--NCRFRP 457
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 509 NLNKR-----VDGSAKWDPYGPCSRTC--GGVQLARRQCTN-- 545
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 458 ARTECRKIDDCDPEYCTGQSECPDVFQRNGQPYQSNNGCYNGNCPILKNQCIHLW 517
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 546 -PTPANG-----GKCYEGVRVYKRSNLEPCPSSASGSKSF 579
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 518 KPAPPAGVNVAPDVCFDNQNGTDVYVNCGIKNGTYIKCARQDIK--CGRLSCEEPSTGNTI 576
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 580 REQCEAFNGYNHSTNRLTLAVAWPKYSGVSPRDCK--LICRAN 623
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 577 ---NCQAFPS---SRN-----PDYGMVDIGTKCADGKVCNSN 607
    || : : : : : || : : : : : || : : : : : || : : : : : || :
```

RESULT 15

S48169
N:Contains: disintegrin
C:Species: Echis pyramidum leakeyi
C:Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 09-Jun-2000
C:Accession: S48169
R:Palme, M.J.I.; Moura-Da-Silva, A.M.; Theakston, R.D.G.; Crampton, J.M.
Eur. J. Biochem. 224, 483-488, 1994
A:Title: Cloning of metalloproteinase genes in the carpet viper (Echis pyramidum leakeyi).
A:Reference number: S48169; MUID:95010025; PMID:7925363
A:Molecule type: mRNA
A:Residues: 1-549 <PAI>
A:Cross-references: EMBL:X78971; NID:g763094; PIDN:CAA55566.1; PID:g763095
C:Superfamily: mouse meltrin alpha; disintegrin homology

C:Keywords: hydrolase; metalloproteinase; venom; zinc
F:1-138/Domain: propeptide (fragment) #status predicted <PRO>
F:139-549/Product: metalloproteinase H-II #status predicted <MAT>
F:350-432/Domain: disintegrin homology <DIS>
F:284/Active site: Glu #status predicted

```
Query Match      8.4%; Score 313; DB 2; Length 549;
Best Local Similarity 23.7%; Pred. No. 1.5e-14;
Matches 156; Conservative 77; Mismatches 234; Indels 192; Gaps 35;

Qy 62 EDFVLHLTPDAQFLAPAFSTEHLGVLPGLGGSDLRRCFYGDVNAEPDSFAAVSLQG 121
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 12 EPVVLHQLKNDLFSEDIYSETHYTPDQGREITNPPVEDHCYIYHGRVQNDHAHSASISACN 71
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 122 GLRGAFGYRGAEEYVISPL--PNASAPAA-----QRNSQGAHLRQRRVPGSPGDPTS 173
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 72 GLKGHFKLOGETYFIEPLKIPDSEAHAYKYENIEKEDQAPK----- 114
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 174 CGVA--SGWNPAILRALDPYKPRRAGFSGESRRSRGRAKRFVSIPIRYVETLVVADESVMK 232
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 115 CGVTHTNW-----ESDEPIK-----EASRLVASSEQQSYIDNFRYVKLVIVVDHRMVT 162
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 233 FHGAD--LEHYLLTLATAARLYRHPISILPINIVVVKVLLLRDRDSGPKVTGNAALT 290
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 163 KYNDSTIRTRYEMVNTVNEIYH---LN--IRVALGLEFWSNGDL--INVTSAEHTLN 217
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 291 NFCAWQKLNKVS-----RHPEYWDTAILFTRODLGGATTCDTLGMADYGTMDPKRSC 345
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 218 LFGVW-----RASLLSRRRH---DNAHLTLAIDLNG---PTIGLRDYSMCAQATRSV 264
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 346 SVIEDDGLPS-----AFTAHELGHVFNPHDNVNVKCEVEFGKLRANHMSPTLIQIDRAN 401
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 265 GWVQDHS--PTVRVAVTMAHEMGNLGNSHDG--NHC-----NCGANSCIMAAVLRNPAPE 317
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 402 PWSACSAIITDFLDGSHGDCLLDQPSK--PISLPEDLPASVYTLSSQOCELAFGVGSKPCP 460
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 318 YFDCSRRYQNFNTNIPDCTIIRPSKTDIVSPQVCGNGLLDEGECD----- 366
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 461 YMOYCTKLWCTGKAGQWVCQTRHFPWADGTSCGEGKCLKGACVERHNLNKHVRDGSWA 520
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 367 -----CGSPAN-----CQ---YPCDDAASC--KL-----HSWV 389
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 521 KWDYGPGRSRTCGGVQVLARQCTNPTPANGGKCEGVVRVYKRSNLEPCPSSASGSKSF 580
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 390 ECE--FGHCCDQC-----RFPKA--GTECRGIR---SECDDL---PEYCTGQS-- 426
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 581 EEQCEAFNGYNHSTNRLTLAVAW---VPKY-----SGVSPRDKC----- 616
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 427 AECLMEFSTRWKTWPKLSLLQLWTCPIWQCYAHFGQNAVVGQDACEFINKKGGKDFYC 486
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Qy 617 -----KLICRA--NGTGYFYVLA PKVVDGTLCSPDSTSVGVQKCI 655
    || : : : : : || : : : : : || : : : : : || : : : : : || :
Db 487 RKENDVPIPCAQEDVKCGRLFCETEPNNCRYPYDGEWMDPGTKC--EDKKVCINGKCI 543
    || : : : : : || : : : : : || : : : : : || : : : : : || :
```

Search completed: April 29, 2003, 17:13:30
Job time : 20.8904 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 8.4611 Seconds
(without alignments)
3367.676 Million.cell updates/sec

Title: US-10-009-332-1_COPY_1_687

Perfect score: 3744

Sequence: 1 MLLGLITLAFAGTAGGFE.....RFDKGVCGDNKCKKYTG 687

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1960.5	52.4	967	1	ATSL_HUMAN	Q9u18	homo sapien
2	1947.5	52.0	968	1	ATSL_MOUSE	P37857	mus musculus
3	1934	51.7	967	1	ATSL_RAT	Q9wu1	rattus norv
4	1639.5	45.4	890	1	ATSL_HUMAN	Q9up9	homo sapien
5	1689.5	45.1	837	1	ATSL_HUMAN	Q5173	homo sapien
6	1679	44.8	905	1	ATSL_MOUSE	P57110	mus musculus
7	1613	43.1	930	1	ATSL_MOUSE	Q9r001	mus musculus
8	1587.5	42.4	930	1	ATSL_HUMAN	Q9una0	homo sapien
9	1543.5	41.2	1629	1	ATSL_HUMAN	Q9p2b4	homo sapien
10	1518.5	40.6	630	1	ATSL_RAT	Q9esp7	rattus norv
11	1087	29.0	997	1	ATSL_HUMAN	Q9uk54	homo sapien
12	1072.5	28.6	1077	1	ATSL_HUMAN	Q9h324	homo sapien
13	1023	27.3	1593	1	ATSL_HUMAN	P58397	homo sapien
14	994	26.5	1211	1	ATSL_HUMAN	Q95450	h adamts-2
15	980.5	26.2	1205	1	ATSL_BOVIN	P79331	b adamts-2
16	968.5	25.9	1205	1	ATSL_HUMAN	O15072	homo sapien
17	918.5	24.5	860	1	ATSL_HUMAN	Q9ukp5	homo sapien
18	601.5	16.1	245	1	ATSL_BOVIN	Q9tt93	bos taurus
19	599	16.0	207	1	ATSL_BOVIN	Q9tt92	bos taurus
20	366.5	9.8	824	1	AD08_HUMAN	P78325	homo sapien
21	360.5	9.6	776	1	AD28_MACFA	Q9xsl6	macaca fasc
22	346.5	9.3	813	1	AD33_HUMAN	Q9bz11	homo sapien
23	344.5	9.2	571	1	DI5J_BOTJA	P30431	bothrops ja
24	342.5	9.1	774	1	AD28_MOUSE	Q9iln6	mus musculus
25	336.5	9.0	775	1	AD28_HUMAN	Q9ukq2	homo sapien
26	331.5	8.9	956	1	AD19_HUMAN	Q9h013	homo sapien
27	329	8.8	903	1	AD12_MOUSE	Q61824	mus musculus
28	325.5	8.7	920	1	AD19_MOUSE	Q35674	mus musculus
29	323	8.6	857	1	AD22_MOUSE	Q9rlv6	mus musculus
30	315.5	8.4	906	1	AD22_HUMAN	Q9p0k1	homo sapien
31	314	8.4	909	1	AD12_HUMAN	Q43184	homo sapien
32	313	8.4	826	1	AD08_MOUSE	Q05910	mus musculus
33	307.5	8.2	789	1	AD07_RAT	Q63180	rattus norv

RESULT 1

ID	ATSL_HUMAN	STANDARD	PRT	967 AA
AC	Q9UH18; Q9UP80; Q9UH83; Q9P2K0; Q9NSJ8;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).			
GN	ADAMTS1 OR METH1 OR KIAA1346.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Casas C., Pritchard M.A., Estivill X., Arbones M.L.;			
RT	"Cloning, characterization and mapping on human chromosome 21 of the orthologue of murine Adamts-1."			
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A., AND FUNCTION.			
RX	TISSUE=Heart;			
RX	MEDLINE=99367466; PubMed=10438512;			
RA	Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,			
RA	Lombardo M., Iruela-Arispe M.L.;			
RT	"METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity."			
RL	J. Biol. Chem. 274:23349-23357(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	TISSUE=Endothelial cells;			
RX	MEDLINE=20247184; PubMed=10785405;			
RA	Glienke J., Schmitt A.O., Pillarsky C., Hinzmann B., Weiss B.,			
RA	Rosenthal A., Thierauch K.H.;			
RT	"Differential gene expression by endothelial cells in distinct angiogenic states."			
RL	Eur. J. Biochem. 267:2820-2830(2000).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	TISSUE=Brain;			
RX	MEDLINE=20181126; PubMed=10718198;			
RA	Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;			
RT	"Prediction of the coding sequences of unidentified human genes.XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro."			
RL	DNA Res. 7:65-73(2000).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20289799; PubMed=10830953;			
RA	Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,			
RA	Park H.-S., Fujiyama A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,			
RA	Ohki M., Takaki T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,			
RA	Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,			
RA	Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,			
RA	Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,			

P15503 trimeresuru
Q9pw35 agkistrodon
O35227 mus musculus
P15167 crotalus at
O75078 homo sapien
Q13444 homo sapien
P34182 crotalus at
Q9rlv4 mus musculus
Q923w9 mus musculus
Q9h2u9 homo sapien
O42596 xenopus lae
Q28476 macaca fasc

ALIGNMENTS

34 297 7.9 480 1 DISA_TRIGA
35 296 7.9 413 1 ACLA_ACKAC
36 288 7.7 788 1 AD07_MOUSE
37 284 7.6 414 1 HRTD_CROAT
38 279.5 7.5 769 1 AD11_HUMAN
39 277.5 7.4 814 1 AD15_HUMAN
40 273.5 7.3 478 1 HRTE_CROAT
41 270.5 7.2 773 1 AD11_MOUSE
42 268 7.2 797 1 AD33_MOUSE
43 260.5 7.0 754 1 AD07_HUMAN
44 260.5 7.0 935 1 AD22_XENLA
45 257 6.9 776 1 AD07_MACFA

RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramer J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Yaspino M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [6]
RN SEQUENCE OF 418-967 FROM N.A.
RC TISSUE-Melanoma;
RA Blum H., Bauersachs S., Mewes H.-W., Weill B., Wiemann S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER (BY SIMILARITY). HAS ANGIOGENIC INHIBITOR
CC ACTIVITY. ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE.
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1938-GLU-|-LEU-1939
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF170084; AAF15317.1; -
DR EMBL; AF060152; AAD48080.1; ALT_INIT.
DR EMBL; AF207664; AAF23772.1; -
DR EMBL; AB037767; BAA92584.1; ALT_INIT.
DR EMBL; AP001697; BAA95502.1; -
DR EMBL; AL162080; CAB82413.1; -
DR MEROPS; M12.222; -
DR Genew; HGNC:217; ADAMTS1.
DR MIM; 605174; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTPeptidse.
DR Pfam; PF00090; tsp_1; 6
DR Pfam; PF01421; Reprolysin; 2.
DR Pfam; PF01562; Pep_M12B_propep; 2.
DR SMART; SM00209; TSP1; 3.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS00092; TSP1; 2.
DR PROSITE; PS00143; ZINC_PROTEASE; 1.
KW Hydrolyase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 49
FT PROPEP 50 252 BY SIMILARITY.
FT CHAIN 253 967 ADAMTS-1.
FT SITE 198 198 CYSTEINE SWITCH (POTENTIAL).
FT METAL 401 401 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 402 402 ZINC (BY SIMILARITY).
FT METAL 405 405 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 411 411 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 476 559 DISINTEGRIN-LIKE.
FT DOMAIN 560 616 TSP TYPE-1 1.

FT	DOMAIN	617	724	CYS-RICH.
FT	SPACER.	725	849	
FT	DOMAIN	850	908	TSP TYPE-1 2.
FT	DOMAIN	909	967	TSP TYPE-1 3.
FT	DOMAIN	843	846	POLY-LYS.
FT	CARBOHYD	547	547	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	720	720	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	764	764	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	227	227	P -> A (IN REF. 4 AND 5).
FT	CONFLICT	468	468	Q -> H (IN REF. 1).
FT	CONFLICT	561	561	S -> N (IN REF. 1).
SEQ	SEQUENCE	967 AA;	105383 MW;	C189389324741ED1 CRC64;
Query Match 52.4%; Score 1960.5; DB 1; Length 967;				
Best Local Similarity 53.6%; Pred. No. 2.9e-138;				
Matches 388; Conservative 91; Mismatches 178; Indels 67; Gaps 17;				
QY	1	MLLIGILTLAGRTAGGFEPEREVVPIRLDPDINGRRYVWRGPDSDGQGLLFOITAF	60	
DB	36	LLLLAAALLAVSDALGRPSEDEELVP-ELE-----RAP---GHGTTTLRLHAF	81	
QY	61	QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRFCYFSGDVNAEPDSF	114	
DB	82	DQDLLELRPDSSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHGFCYSGVNGDPSSA	136	
QY	115	AAVSLCGGLRGAFYVGAAYVISPLPNAS---APAAQNSOGA---HLLQ--RRGVPGG	165	
DB	137	AAALSCEGVGAFYLLGEAYFIQPLPAASERLATAAPGKPPALQPHLLLRNRQGVGG	196	
QY	166	PSG-----DPTSRC-----GVASG--WNPAILRALDPYKPRRAGFSGSRRRS	207	
DB	197	TCGVVDDPRFTGKAEDEDEGEDEGPGQWS-----PQDPALQGVGP--TGTS	247	
QY	208	GRAKREVSIPRYVETLVVADSSMKFAGDALEHYLLTLATAARLYRHPSTILNIVVV	267	
DB	248	IRKKRFFVSHRYVETMLVADSSMAEFHSGLGKHYLLTLFSAARLYRHPSTILNIVVV	307	
QY	268	KVLLLRDRDSGPKVTGNAALTRNFCAMQKLNKYSKDHPEWDTAIFTTQDLCGATTC	327	
DB	308	KILVTHDEOKGPEVTSNAALTURNFCNQKQHNPPSDRDAEHYDTAIFTTQDLCGATTC	367	
QY	328	DTLGMADVGTMDCKRSCSVIEDDGLPSAFTTAHELGHVFNHNPVKNVKEVFGKLRAN	387	
DB	368	DTLGMADVGTVCDSRSCSVIEDDGLQAFTTAHELGHVFNHNPVKNVKEVFGKLRAN	427	
QY	388	HMSPTLQIDRANPWSACSAIITDFDLSHGDCLLQPSKPSLPEDLPGASVTLSSQ	447	
DB	428	HMASMLSLNLDHSQWSPSCSAYMITSFLDNHGHGCLMDKQNPITQPLGDLPTSTDAHQ	487	
QY	448	CELAFCVGSKPCP-YMQYCTKLWCTGKAKQMVQCOTRIHFPWADGTCGEGKLCAGC	506	
DB	488	COPTFGEDSKHCPDAASTCTLWCTGTSGLVLCOTKHFPAWADGTCGEGKLCAGC	547	
QY	507	RHNLANKH---RYDGSWAKWDYPCGSRCTCGGVQVQARQCTNPTPANGKCYEGVRKYR	563	
DB	548	KTD-RKHFDTPFHGSMGMWPGDCSRTCGGVQVQARQCTNPTPANGKCYEGVRKYR	606	
QY	564	SCNLEPCPSASCSPREQCEAFNCSNLTNRLTAVAWPKYSGVSPDRCKKLCICRAN	623	
DB	607	SCNLEPCPDN-NGKTPREQCEAHNEFSKASGSGPAPWEIYKAGVSPDRCKKLCICRAN	665	
QY	624	GTGYFVLPKPVVDGTLCSPTDSVVCVQCKICAKGCDGNLGSKKRPFKCGVCGGNKSK	683	
DB	666	GIGYFVLPKPVVDGTLCSPTDSVVCVQCKICAKGCDRIIDSKKPKDKGCGVCGGNSTCK	725	
QY	684	KVTS 687		
DB	726	KISS 729		
RESULT 2				
ATSL_MOUSE STANDARD; PRT: 968 AA.				
ID	ATSL_MOUSE			


```
QY 207 -----SGRAKREYIPRYVETLVVADESMVKFHGADLEHYLLTLLATAARLYRHPHSILNPI 262
DQ 244 SGPFSIRKKREYSSPYVETMLVADQSMADFHGSLGKHYLLTLESVAARYKHPSIRNSI 303
QY 263 NIYVVKVLLLRDSDSPKVTGNAALTLRNFCAWQKLNKVSQKHPYWDYDAILFTRQDLC 322
DQ 304 SLVYVVKILYIEQKGPVTSNAALTLRNFCAWQKLNKVSQKHPYWDYDAILFTRQDLC 363
QY 323 GATTCDTLGMADVGTWCDPKRSCSVTEDDGLPSAFTTAHELGHVFNMPHDNVKVCCEVFG 382
DQ 364 GSHTCDTLGMADVGTWCDPKRSCSVTEDDGLPSAFTTAHELGHVFNMPHDNVKVCCEVFG 423
QY 383 KLRANHMSPTLIQIDRANPWSACSAIITDFLDHSGHGLDQSPKISLPEDLPAGSY 442
DQ 424 VTGDSHLMASMLSSLDHSPQSPVTSYMTSFLDNGHGBCLMDKPNPIKLPDLPGLTY 483
QY 443 TLSQCELAFLGVSQKCP-PMQYCTKLWCTGKAGKQWQVCTHFWADGTSGBEGKLCIK 501
DQ 484 DANRQCOFTFGEESKCPDAASTCTTLWCTGTSGGLLVQCTKHFWADGTSGBEGKWCVS 543
QY 502 GACVERHNLNKH---RVDGSAWAKWDPYGPCSRTCGGGVQLARRQCTNPTTPANGKGYCEGV 558
DQ 544 GKCVNKTDM-KHPATPVHSGWGPWGPWGPWGPWGPWGPWGPWGPWGPWGPWGPWGPWGP 602
QY 559 RVKYRCNLEPCPSSASGSKSFRECEAFNGYHNHSTNRLTLAVWVPKYSVSPRDKCL 618
DQ 603 RVKYRCNLEPCPSSASGSKSFRECEAFNGYHNHSTNRLTLAVWVPKYSVSPRDKCL 661
QY 619 ICRANGTGVFYLAPKVPVDTGLSPDSTSVCGVQKIKAGCDGNLGGSKKRFKDCGVC 678
DQ 662 TCBAKIGIFVFLQPKVDTGLSPDSTSVCGVQKIKAGCDGNLGGSKKRFKDCGVC 721
QY 679 NKSCCKKVTG 687
DQ 722 GSTCKKMSG 730

RESULT 3
ATSI_RAT
ID ATSI_RAT STANDARD; PRT; 967 AA.
AC Q9WU01; Q9ERIL;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (BC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN ADAMTS1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RA Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,
RA Little S.P.;
RT "Induction of a disintegrin and metalloproteinase with the
RT thrombospondin type I motif (ADAMTS).";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 18-967 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=20304099; PubMed=10847486;
RA Diamantis I., Luethi M., Hoesli M., Reichen J.;
RT "Cloning of the rat ADAMTS-1 gene and its down regulation in
RT endothelial cells in cirrhotic rats.";
RL Liver 20:165-172(2000).
CC -!- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC SIMILARITY). ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
```

-!- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU-1-LEU-1684 SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.

-!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

-!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (by similarity).

-!- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM CIRRHOTIC LIVER.

-!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

-!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

-!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

-!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF149118; AAD34012.1; -

EMBL; AF304446; AAG29823.1; -

MEPROS; M12.222; -

InterPro; IPR001762; Disintegrin.

InterPro; IPR002870; Rep_M12B_propep.

InterPro; IPR001590; Repolysin.

InterPro; IPR000884; TSP1.

InterPro; IPR000130; Zn_MTPeptidse.

Pfam; PF00090; tsp_1; 6.

Pfam; PF01421; Repolysin; 2.

Pfam; PF01562; Pep_M12B_propep; 2.

SMART; SM00209; TSP1; 3.

PROSITE; PS02015; ADAM_MEPRO; 1.

PROSITE; PS00142; ZINC_PROTEASE; 1.

PROSITE; PS00092; TSP1; 2.

PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.

Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen; Repeat; Extracellular matrix; Heparin-binding.

POTENTIAL.

SIGNAL 1 54

PROPEP 55 252

CHAIN 253 967

FT SITE 205 205

FT METAL 401 401

FT ACT_SITE 402 402

FT METAL 405 405

FT METAL 411 411

FT METAL 476 558

FT DOMAIN 559 615

FT DOMAIN 616 724

FT DOMAIN 725 857

FT DOMAIN 858 907

FT DOMAIN 908 967

FT DOMAIN 194 198

FT CARBOHYD 547 547

FT CARBOHYD 720 720

FT CARBOHYD 764 764

FT CARBOHYD 782 782

FT CARBOHYD 945 945

FT CONFLICT 21 21

FT CONFLICT 26 31

FT CONFLICT 49 49

FT CONFLICT 72 72

FT CONFLICT 79 79

FT CONFLICT 249 249

FT CONFLICT 262 265

FT CONFLICT 607 607

FT CONFLICT 936 936

FT CONFLICT 962 962

FT CONFLICT 967 967

SEQUENCE 105705 MW; F93C864F6DCDB4CF CRC64;


```
Query Match          51.7%; Score 1934; DB 1; Length 967;
Best Local Similarity 52.7%; Pred. NO. 2.8e-136;
Matches 378; Conservative 98; Mismatches 187; Indels 54; Gaps 15;

QY 1 MLLGLTILAFAGRTAGG--FPEREVVVPIRLDPDINGRRYYRWGPEDSGDGLIFQIT 58
DB 1 LLLASITMLLCVGAHGRPTDEDELVL-----PSLRARGH-----DSTVLLRLD 83
QY 59 AFQEDFVHLTDAFAPAFSTELHGLVPLGLTGG-----SDLRCFYSGDVNAE 110
DB 84 AFQQLHLKLQDPSGLFAPGFTLQV-----GRSPGSAQHLDPDGLAHCFYSYVNGD 138
QY 111 PDSFAVSLCGGLRGAFGYRGAEYVISPILPNAS-----APAAQRNSQGA-----HLL--QRR 160
DB 139 PSSAALSCEGVGAFYLGEEFTIQAPAVATERLVPAPKEESIAPRFHILRRRR 198
QY 161 GVPGPSGD-----PTSECGVASGWNPAILLRALDPYPRAGFESRRRSRAKRFV 214
DB 199 GGGAKCGVMEETLFTNSGRESQNTPDWPLRNP--TPQAG-----KPTGGSIRKKRFV 254
QY 215 SIPRVETLVADSMVRFHGLADLEHLLTLLATARLYRHPISILNPINIVVKKVLLRD 274
DB 255 SPSRVETLVADSMVRFHGLADLEHLLTLLFSAARYKHPISIRNSISLVKKVLIYIE 314
QY 275 RDSGPKVTGNAALTNRFCAMOKLNKYSDKHPEYWDTAILFTRODLGCATTCDTLGMD 334
DB 315 EQGPEVTSNAALTNRFCAMOKLNKYSDKHPEYWDTAILFTRODLGCATTCDTLGMD 374
QY 335 VGTMDCKPRSCSVIEDGLPSAFTAHGLGHVFNPHNVKVCVEVFKGLRANHMSP 394
DB 375 VGTVCDPSCSVIEDGLPSAFTAHGLGHVFNPHNVKVCVEVFKGLRANHMSP 434
QY 395 IOIDRANPWSACSAIITDLSHGDCILLOPKSPISLPEDLPACATYLSQCELAGV 454
DB 435 SLDSHSGWSPSCSAVYVTFSLDNGHECLMDKPNQPIKLPDLPTLYDANRQCOFTGE 494
QY 455 GSKPCP-YMQYCTKWCCTGKAKGMVQCOTRHPFWADGTSCGEGKLCGLKACVERHNLKH 513
DB 495 ESTHCPDAASTCTLWCTGTSGLLVLCOTKHPWADGTSCGEGKWCVCVKVNTDM-KH 553
QY 514 ---RVDSWAKWDPYGPSCRTCGGVQALROCTNPTFANGKCYCEGVVRYRSCNLEPC 570
DB 554 FATPVHSGWPGWPGDCSRTCGGVQVYTMRECDNPVPKNGKCYCEGVVRYRSCNLEPC 613
QY 571 PSSAGKSPREOCFAFNGYHNSTNRLTLAVAWPKYSGVSPDKCKLICRANGTGYEV 630
DB 614 PGN-NKGFREOCFAHNEFSASGNEPTVETPKYAGVSPDKCKLICRANGTGYEV 672
QY 631 LAPKVVDTLSPDSTVSCVQKCIKAGCDNGLGSKRFRDGGVCGGDNKSKCKVTG 687
DB 673 LQPKVVDTLSPDSTVSCVQKCIKAGCDNGLGSKRFRDGGVCGGDNKSKCKVTG 729

RESULT 4
AT58_HUMAN          STANDARD;          PRT;          890 AA.
AC Q9UP79; Q9NZS0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2)
DE (METH-8).
GN ADAMTS8 OR METH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=99367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
```

```
RA Lombardo M., Iruela-Arispe M.L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
RT family of proteins with angio-inhibitory activity.",
RL J. Biol. Chem. 274:23349-23357(1999).
RN [2]
RP SEQUENCE OF 195-440 FROM N.A.
RX MEDLINE=20079168; PubMed=10610729;
RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on
RT mouse chromosome 9 and human chromosome 11.",
RL Genomics 62:312-315(1999).
CC -1- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER
CC KIDNEY.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF060153; AAD48081.1; -
DR EMBL; AF175283; AAP25806.1; -
DR HSSP; P34179; IIAQ.
DR MEROPS; M12.226; -
DR Genew; HGNC:224; ADAMTS8.
DR MIM; 605175; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reptolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTPpeptdse.
DR Pfam; PF000090; tsp_1; 2.
DR Pfam; PF01421; Reptolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 2.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS50092; TSP1; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL; 1 27 POTENTIAL.
FT PROPEP; 28 214 BY SIMILARITY.
FT CHAIN; 215 890 ADAMTS-8.
FT METAL; 364 364 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE; 365 365 BY SIMILARITY.
FT METAL; 368 368 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL; 374 374 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN; 439 526 DISINTEGRIN-LIKE.
FT DOMAIN; 527 583 TSP TYPE-1 1.
FT DOMAIN; 584 690 SPACER.
FT DOMAIN; 691 832 CYS-RICH.
FT DOMAIN; 833 890 TSP TYPE-1 2.
FT DOMAIN; 202 205 POLY-PRO.
FT CARBOHYD; 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD; 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD; 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD; 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD; 600 600 N-LINKED (GLCNAC. . .) (POTENTIAL).
```



```
DR MM; 603876;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTPeptide.
DR Pfam; PF00090; tsp_1; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR SMART; SM00209; TSPl; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS0092; TSPl; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Extracellular matrix.
FT SIGNAL 1 51 POTENTIAL.
FT PROPEP 52 212
FT CHAIN 213 837
FT SITE 194 194
FT METAL 361 361
FT ACT_SITE 362 362
FT METAL 365 365
FT METAL 371 371
FT DOMAIN 437 519
FT DOMAIN 520 576
FT DOMAIN 577 685
FT DOMAIN 686 837
FT DOMAIN 247 252
FT CARBOHYD 68 68
FT CONFLICT 77 77
FT CONFLICT 626 626
FT CONFLICT 682 682
FT SEQUENCE 837 AA; 90224 MW; 5DF9C9AC137DF41F CRC64;
Query Match
Best Local Similarity 45.1%; Score 1689.5; DB 1; Length 837;
Matches 328; Conservative 103; Mismatches 215; Indels 49; Gaps 14;
QY 1 MLLGILLIATAGRTAGGFEFEEVVPRLDPDINGRRYYWRGPDSDGDLGFIQTAF 60
DB 37 LLLLALLSLLSARLASPLPEETIEVFEKLNGSVL-----PGSGAPARLLCLRQAF 88
QY 61 QEDFVHLTPDPAFLAPAFSTHGLVPLQGLTGSSDLRRCFYSGDVNAEPDPSFAAAYSLC 120
DB 89 GTLLLELEQDSGVQVEGLTVQLQAPF--LLGAEF--GTYLTGTNDGPESVASLHWD 145
QY 121 GG-LRGAFGYGAEVYISPLPNASAPAAQRNSQGAHLQRRVPGSPGSDPTSCRGVASG 179
DB 146 GCALLGLVQYRGAELHLQPLEGGTPNSA--GGPGAHLRRK-----SPASGGGPMCNV--- 196
QY 180 WNPATLRALDPYKPRAGFGESRRSRGRKRFRVSPRYVETLVVADESMVKFHGADLE 239
DB 197 -----KAPLGSPSPRPR--RAKRFASLSRFVETLVVADKMAAFHAGLKL 239
QY 240 HYLTLTLLAARLYRHPISILNPINIVVVKVLLRRDQSGPKVTCGNAALTLRNFCAWQKL 299
DB 240 RYLLTVMAAAKAKHPISIRNPVSLVTRVILGSGEGPGVPGPSAAQTLRSCAWQRL 299
QY 300 NKVSDKHPYNDTAILFTRODLGATTCDTLGMADVGTMCDPKRSCSVIEDGLPSAFTT 359
DB 300 NTPEDSDPDHFDTAILFTRODLGATTCDTLGMADVGTMCDPKARSCAIVEDDGLQSAFTA 359
QY 360 AHELGHVFNMPHDNVKVEEYFGKL--RANHMMSTLQIDRANPWSACSAITDFDLDSG 418
DB 360 AHELGHVFNMLHDSKPCISLNGPLSTSRHYMAPVMAHVDPPEEPWSPCSARFITDFDLNG 419
QY 419 HGDCILLDPKSPISLPEDLPASVTLTSSQCELAFGVSKGPCPYM--QYCTKLWCTGKAKQ 477
DB 420 YGHCLLDKPEAPLHLPTVTFKQDADRCQCLTEGPDNRHCPQLPPCAALWCSGHLNGH 479
QY 478 MYCOTRHPFWADGTSCEGKGLCLGACVERHNLNKRVD--GSWAKWDYPGCSRTCCGG 535
DB 480 AMCQTKHSPWADGTPCGPAQACMGRCGLHMDQLQDENIPQAGGWPWGPGWDCSRTCCGG 539
```

```
QY 536 VOLARROCTNPTPANGKYCEGVKVRKYNLCNLEPCPSSASGSKSPREBQCEAFNGYNHSTN 595
DB 540 VQFSRCDCTRPVPRNGKYCEGRTRFRSCNTEDCP--TGSALTFREEQCAA---YNHRTD 595
QY 596 RLTL---AVAWVPKYSGVSPDKCKLICRANGTGYFYVLAPKVVVDGTLCSPDSTSVQVQ 652
DB 596 LFKSFPQMDWVPRTYGVAPQDQCKLTQARALGYVYVLEPRVVDGTPCSPDSSVQVQ 655
QY 653 KCIAGCDGNLGSKKRFDKCGVCGGDNKSKKVTG 687
DB 656 RCHAGCDRTIGSKKFKDKMCGGDSGSGSKQSG 690
RESULT 6
AT88_MOUSE
ID AT88_MOUSE STANDARD; PRT; 905 AA.
AC P57110;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2).
GN ADAMTS8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RX SEQUENCE FROM N.A.
RP MEDLINE=20079168; PubMed=10610729;
RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on
mouse chromosome 9 and human chromosome 11.";
RL Genomics 62:312-315(1999).
CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES (BY SIMILARITY).
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT LUNG AND HEART
AND LOW EXPRESSION DURING MOUSE DEVELOPMENT.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF175282; IAAF25805.1;
DR HSP; P34179; IIAF;
DR MEROPS; M12.226;
DR MGD; MGI:1353468; Adamts8.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTPeptide.
DR Pfam; PF00090; tsp_1; 2.
DR Pfam; PF01421; Reprolysin; 1.
DR SMART; SM00209; TSPl; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS0092; TSPl; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
```

FT	SIGNAL	1	28	POTENTIAL.
FT	PROPEP	29	228	BY SIMILARITY.
FT	CHAIN	229	905	ADAMTS-8.
FT	METAL	378	378	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	379	379	BY SIMILARITY.
FT	METAL	382	382	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	388	388	ZINC (CATALYTIC) (BY SIMILARITY).
FT	DOMAIN	453	541	DISINTEGRIN-LIKE.
FT	DOMAIN	542	598	TSP TYPE-1 1.
FT	DOMAIN	599	705	CYS-RICH.
FT	DOMAIN	706	847	SPACER.
FT	DOMAIN	848	905	TSP TYPE-1 2.
FT	CARBOHYD	415	415	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	480	480	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	506	506	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	615	615	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE	905 AA;	98879 MR;	124D4132B33A0CAE CRC64;
Query Match 44.8%; Score 1679; DB 1; Length 905;				
Best Local Similarity 46.7%; Pred. No. 2.4e-117;				
Matches 344; Conservative 101; Mismatches 202; Indels 90; Gaps 18;				
Qy	2	LLLGILRL-----AFAGRTAGGPEPEVVPVPIRLDPDINGRYVWRGPEDSGDQGL	53	
Db	13	LLLLLLPPPLVCGAPAGGTGA--QASELVVPTRL-----PGSASE--L	55	
Qy	54	IFQITAFQEDFYHLTDAQFLAPAFSTEHLGVPLQGLTGGSDLRRCFYSGDVNAEPDS	113	
Db	56	AFLSARGQGVFLRADFASFLAPEFKIERLG-GSSAAAGGEPGLRCGPFSGVNGERES	114	
Qy	114	FAAYSCGGLRGAGYGAEVYISPLNAPAAQNSQGA-----HLLQRGVPGGP	166	
Db	115	LAAMSCVAGSGSPLLAGEEFTIQP-----QAGDSLDPHRLQRWG-PGQR	160	
Qy	167	SGDPTSCRGVAGSNPAILRALDPYKPRRAGFGEERS-----RRSG-----	208	
Db	161	REDP---GLAAAEVFLPGQLEWEVEMNGOQOERDNEEDRKQDEGLLKETEDSRKV	216	
Qy	209	-----RAKRFYSIPRYVETLVVADPMVKFGADLEHLYLTALTAARLYRHPISLN	260	
Db	217	PPPGSKTRSRFFSEARFVETLLVADASMAAFVGTDLQNHILTVMSAARIYKHPISRN	276	
Qy	261	PINIVVKVLLLRDRSDQKVTGNAALTLRNFCAMQKLLKNSDKHPYWDITAILFTRQD	320	
Db	277	SVNLVVKVLLIVERKGPVSDNGGLTLRNFCSQWRFNKPSDRHPHYDHTAILFTRQN	336	
Qy	321	LCG-ATTCDTLGMADVGTCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKKVCEE	379	
Db	337	FCGKGEQCDTLGMADVGTCDPKSCSVIKDEGLQAAYTLAELHGHVLSMPHDDSKPCVR	396	
Qy	380	VFGKLRANHMSPTLIQIDRANPWSACSAAITDFLDSGHGDCILLDQPSKPISLPDLPG	439	
Db	397	LFGPMGYHMAFFFIHVNTLPWSPCSAVYLTLLDGDHGDCLLDAPTSVLPLTGLPG	456	
Qy	440	AS--YTLSSQCELAFGVSKPCP---TMOYCTKLWCTGKAGQWVCOTRH--FPWADGTS	492	
Db	457	HSTLYELDQCKQIFGPDFRCPNTSVEDICVQLCARHRSDEPICHTKNGSLWADGTP	516	
Qy	493	CGEGLCLKLGACVBRHNLN--KHRVDGWAKWDYPGCSRTCGGVQLARRQCTNPTPAN	550	
Db	517	CGPHLCLDGSCLVKEDVENPKAVVDGWDGWPFWGQCSRTCGGGIQFSNRECDNPNPQN	576	
Qy	551	GKTYCEGVRYKRYSCNLPCCPSSASGSKSFREEQCEAFNGYHNHSTNRLTLAVAWPKYSGV	610	
Db	577	GGRFCLGERVYQSONTEECF--PNGSKFRQCEKYNAYNH--TDLGDNFLQWPKYSGV	633	
Qy	611	SPRDKCLICRANGTYFYVLAPKVVDTLCSPDSTSVCVQKCIKACGCGNKKSKREFD	670	
Db	634	SPDRCKLFCRARGSEKFEAKVIDGTLCGPDTLSTICVRGQCVKAGCDHVNPSPKLKD	693	
Qy	671	KGVCGGDNKSKKVTG 687		
Db	694	KGVCGGKGTACKRTSG 710		

RESULT 7				
ID	ATSS_MOUSE	STANDARD;	PRT;	930 AA.
AC	09R001;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2) (ADMP-2) (Implantin).			
DE	ADAMTS5.			
GN	Mus musculus (Mouse).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A. PubMed=10464288;			
RX	MEDLINE=99395124; PubMed=10464288;			
RA	Hurskainen T.L., Hirohata S., Seidin M.F., Apte S.S.;			
RT	"ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of			
RT	zinc metalloproteinases.";			
RT	J. Biol. Chem. 274:25555-25563(1999).			
CC	FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.			
CC	CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393 site.			
CC	COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).			
CC	SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).			
CC	DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PERI-IMPLANTATION PERIOD IN EMBRYO AND TROPHOBLAST AND AT LOW OR UNDETECTABLE LEVEL THEREAFTER.			
CC	DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.			
CC	PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).			
CC	SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.			
CC	SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.			
CC	SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
DR	EMBL; AF140673; AAD56356.1;			
DR	MEROPS; M12.225;			
DR	MGI; 1346321; Adamts5.			
DR	InterPro; IPR001762; Disintegrin.			
DR	InterPro; IPR002870; Pep_M12B_propep.			
DR	InterPro; IPR001590; Reprolysin.			
DR	InterPro; IPR000884; TSP1.			
DR	InterPro; IPR00130; zn_MTpeptdse.			
DR	Pfam; PF00090; tsp.1; 2.			
DR	Pfam; PF01421; Reprolysin; 1.			
DR	Pfam; PF01562; Pep_M12B_propep; 1.			
DR	SMART; SM00209; TSP1; 2.			
DR	PROSITE; PS0215; ADAM_MEPRO; 1.			
DR	PROSITE; PS0092; TSP1; 1.			
DR	PROSITE; PS00442; ZINC_PROTEASE; 1.			
DR	PROSITE; PS00427; DISINTEGRIN.1; FALSE NEG.			
KW	hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;			
KW	Repeat; Extracellular matrix.			
FT	SIGNAL	1	21	POTENTIAL.
FT	PROPEP	22	261	
FT	CHAIN	262	930	ADAMTS-5.

```
FT SITE 209 209 CYS-STEINE SWITCH (POTENTIAL).
FT METAL 410 410 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 411 411 BY SIMILARITY.
FT METAL 414 414 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 420 420 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 485 566 DISINTEGRIN-LIKE.
FT DOMAIN 567 623 TSP TYPE-1.
FT DOMAIN 624 731 CYS-RICH.
FT DOMAIN 732 874 SPACER.
FT DOMAIN 875 930 TSP TYPE-1 2.
FT DOMAIN 41 46 POLY-ALA.
FT DOMAIN 257 261 POLY-ARG.
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 728 728 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 930 AA; 101780 MW; 84DE84B26170D4DC CRC64;

Query Match 43.1%; Score 1613; DB 1; Length 930;
Best Local Similarity 47.4%; Pred. No. 2,1e-112;
Matches 306; Conservative 81; Mismatches 182; Indels 76; Gaps 7;

QY 90 GLTGSSDLRCFYSGDVNAEPDFAVSLCGGLRGAFYRGAEYVISPFPNASAPAAQR 149
DB 121 GLSASSHGRHCFYGTGVDGSPRSLAVFDLCGLDGFVAVKHARYTLKPLLRGSWAVER 180
QY 150 -----NSQAHLLORRG-----VPGPGSGDPTSCRGVAGSNWPAILRALDPY 191
DB 181 IYDGSRLHYNREGFSEFALPRASCETPASPFG----- 217
QY 192 KPRAGFGESESRSSG-----RAKRFVSIPIRYVELTVAD 227
DB 218 -POESPSVHSRRSRSSALAPOLLDSAFSPSGNAGPQTWRRRRRSISRARQVELLVAD 276
QY 228 ESMVKFHGADLEHYLLTLTATAARLYRPSLTLPINIVVKKVLLLRDSDGPKVTGNAAL 287
DB 277 SSMRWYGRGLHLYLLTASTANLYSHASTENHILAVKVVVLTDKDTSLEYSKNAAT 336
QY 288 TLRNFCAMQKLNKVSOKHPEYDWTAILFTQDLGCGATTCDTLGMADVGMCDPKRSCSV 347
DB 337 TLKNFCKWOHQHNGDDHEHYDAALFTREDLCGHHSCTDLGMADVGTICSPERSCAV 396
QY 348 IEDDGLSAFTAHGLHGVFNHNDNVKVEEFGVKLRANHMMSPTLIQIDRANPWSACS 407
DB 397 IEDDGLHAAFTVAEIGHGLSHDSKFCFEEFGTTEDKRLMSILTSIDASKPWSKCT 456
QY 408 AAILTFDLDGSHGCDLLDPSKPLSLPDLPGASYTILSQCELAFGVGSKPCPYMQYCTK 467
DB 457 SATITEFLDGHGNCILLDLPRKQLGPEELPGQYDQTCNLTFGPEYVCPGMDVCAR 516
QY 468 LWCTGKAGQVMQOTRHPFADGTSCGEGKLCCLKGACVER-----HNLNKHRYDGVGSAKW 522
DB 517 LWCAVVRQGMVCLTKLPAVEGTPCGKRGVCLQKGVCKDKTKKYYSTSSH---GNWGSW 573
QY 523 DPYPCGSRCTGGGVQVLAARCTNPTPANGKYCEGVRYKYSRNLNLEPCPSASGKSFREE 582
DB 574 GPWQCSRGSGGVQVFAIRCHNPNAPRNSRGYCTGKRAIYRSGSVTPCP---PNGKSFRE 631
QY 583 QCEAFNGYNHSTRLTLAVAWPKYSGVSPRDKCLCRANGYGFYVLAPKPVVDGTLCS 642
DB 632 QCEAKNGYQDAKGKVTVEWPKYAGVLPADVCKLTCRAKGHYGVVSPKVTDTGTECR 691
QY 643 PDSTSVCVQKCKTAKAGCDNLGSKKRFDKCGVCGGDNKSKCKVTG 687
DB 692 PYSNSVCRVCRVGTGDDGIIGSKLQYDKCGVCGGDNSSCTKIIG 736
```

RESULT 8

```
ATSS_HUMAN
ID ATSS_HUMAN STANDARD; PRT; 930 AA.
AC Q9UNA0; Q9UKP2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
```

15-JUN-2002 (Rel. 41, Last annotation update)
ADAMTS-5 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 5) (ADAM-TS 5) (ADAM-TS5) (Aggrecanase-2) (ADMP-2) (ADAM-TS 11).
ADAMTS5 OR ADMP2 OR ADAMTS11.
Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=99367476; PubMed=10438522;
RA Abbaszade I., Liu R.-Q., Yang F., Rosenfeld S.A., Ross O.H.,
RA Linn J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M.,
RA Wynn R., Duke J.L., George H.J., Hillman M.C. Jr., Murphy K.,
RA Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner R., Nagase H.,
RA Ito Y., Newton R.C., Magolda R.L., Trzaskos J.M., Hollis G.F.,
RA Arner E.C., Burn T.C.;
RT "Cloning and characterization of ADAMTS11, an aggrecanase from the ADAMTS family";
RL J. Biol. Chem. 274:23443-23450(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Anconarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Schafie M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Daggand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Vaspo M.-L.;
RT "The DNA sequence of human chromosome 21";
RL Nature 405:311-319(2000).
RN [3]
RP SEQUENCE OF 413-930 FROM N.A.
RX TISSUE=Fetal brain;
RA Hurskainen T.T., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of zinc metalloproteases";
RL J. Biol. Chem. 274:25555-25563(1999).
CC -!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393 site.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).
CC -!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE, CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN ARTHRITIC PATIENT.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A PURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF261918; AAF89106.1; -
CC EMBL: AB037733; BAA92550.1; -
CC HSP: P15167; IATL.
CC MEROPS: M12.021; -
CC Genew: HGNC:13202; ADAMTS9.
CC MIM: 605421; -
CC InterPro: IPR001762; Disintegrin.
CC InterPro: IPR002870; Pep_M12B_propep.
CC InterPro: IPR001590; Reprolysin.
CC InterPro: IPR000884; TSP1.
CC InterPro: IPR000130; Zn_M12peptidse.
CC Pfam: PF00090; tsp_1; 11.
CC Pfam: PF01421; Reprolysin; 1.
CC Pfam: PF01562; Pep_M12B_propep; 1.
CC SMART: SM00209; TSP1; 12.
CC PROSITE: PS02015; ADAM_MEPRO; 1.
CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE: PS50092; TSP1; 9.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC Hydrolase: Metalloprotease; Zinc; Signal; Glycoprotein; zymogen;
CC Repeat: Extracellular matrix; Alternative splicing.
CC SIGNAL: 1 18
CC PROPEP: 19 287
CC CHAIN: 288 1629
CC DOMAIN: 509 587
CC DOMAIN: 589 642
CC DOMAIN: 645 752
CC DOMAIN: 753 880
CC DOMAIN: 999 1053
CC DOMAIN: 1056 1108
CC DOMAIN: 1111 1156
CC DOMAIN: 1184 1239
CC DOMAIN: 1240 1295
CC DOMAIN: 1332 1383
CC DOMAIN: 1386 1439
CC DOMAIN: 1445 1498
CC DOMAIN: 1501 1554
CC DOMAIN: 1562 1612
CC DOMAIN: 88 96
CC SITE: 223 223
CC METAL: 434 434
CC ACT_SITE: 435 435
CC METAL: 438 438
CC METAL: 444 444
CC CARBOHYD: 112 112
CC CARBOHYD: 135 135
CC CARBOHYD: 271 271
CC CARBOHYD: 749 749
CC CARBOHYD: 840 840
CC CARBOHYD: 1213 1213
CC CARBOHYD: 1267 1267
CC VARSPLIC: 1064 1072
CC
CC VARSPLIC: 1073 1629
CC CONFLICT: 367 367
CC SEQUENCE: 1629 AA; 182649 MW; C1C4CEFF58B8941F CRC64;

Query Match 41.2%; Score 1543.5; DB 1; Length 1629;
Best Local Similarity 42.5%; Pred. No. 6e-107;
Matches 308; Conservative 99; Mismatches 246; Indels 71; Gaps 13;
QY 22 EREVVVPIRLDP-----DINGRRYV-----RGPEDSGGGLFQITAFQ 61
DB 47 EYEIVSIRVNALGEPPTVNHFKRTRRSINATDWPAPASSSSSTSSQAHYRUSAF 106
QY 62 EDFYLHLTPDAQFLAPAFSTEHLGVP-----LQGLTGSSDLRCFYSGDVNABPDSFAAV 117
DB 107 QQEFLNLTANAGIAPLFTVLLGTPOVNTKYSSEAEALKHCFFKGYVNTNSEHTAVI 166
QY 118 SLGGGLRGAGYGAEVISPLNAPAAORNSQAHLLORRGVPGSGDPTSCGVA 177
DB 167 SLGSMGLTFRSHGDGYFIEPLQSMDEQDEDEEQNKPHIYRRSAPQREPFSTGRHACDTS 226
QY 178 SGWNPAILRALDPYKPRRAGFE-----SRSSRRSG 208
DB 227 EHN-----RHSKDKKTKARKWGERINLAGVAALNSGLATEAFSAVGNKTDNTRERTHR 283
QY 209 RAKRFVSIPIRYVETLVVADESMYKFHGADEHYLLTLATAARLYRHPHSILNPINIVVVK 268
DB 284 RTKRFLSYPRFEVLVVDNRNMYSYHGENLQHYILTMSIVASIIYKDPISGNLINIVVN 343
QY 269 VLLLRDRDSGPKVTGNAALTNRNFCWOKKLNKVSDEKHEWDTAILFTRODLGCA-TTC 327
DB 344 LIVIHNEQDGSISFNAQTTLKNFCOMQHSKNSPGGIH---HDTAVALLTRQDTCRAHDKC 400
QY 328 DTLMADVGTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEVEFGKLRAN 387
DB 401 DTGLAELGTICDPYRSCSISEDSGLSTAFTIAHELGHVFNMPHDNDNNKKEE-GYKSPQ 459
QY 388 HMSPTLIQIDRANPWSACSAIITDLSGHGDCLLDQ-SKPISLPEDLPFASVTLQ 446
DB 460 HVMAPTLNEYTNPMWMSKCSKYITEFLDTGYCECLNNEPESRPYPLVQLPGILYNVK 519
QY 447 QCELAFCVSGSKPCPYQWYCTKLWC---TGKAKQWVCOTRHFHWADGTSCGEGKCLKGA 503
DB 520 QCELIFGSGQVCPYMMQCRRLMCNNVNGVHKG---CRTQHTPWADGTCECEGHCKYGF 576
QY 504 CVERHNLNHRVDGSKWAKWDYPGCSRTGGVQLARRQCTNPTPANGKYCEGVKVKYR 563
DB 577 CVPK-EMDVPTDGSWGSWSPGTCSTCGGKIKTAIRECNPEKPKNGKYCVGRMKKF 635
QY 564 SCNLEPCPSASGSKSPREOCEAFNGYHNSTNRLTLAVAWVPKYSVSPDRCKKLCFRAN 623
DB 636 SCNTEPCLKQK--RDFRDEQCAHFDGKHFNINGLLPNVRWVPKYSIGILMKDRCKLCFVA 693
QY 624 GTGYEYVLAPKVVDTGLCPDSTSVCGKCIKAGCDGNGLSKKRFDKCGVGGDNKSK 683
DB 694 GNTAYQLDRVLDGTPCGQDINDICVQGLCRQAGCDHVLNSKARRDKCGVGGDNSSCK 753
QY 684 KVTG 687
DB 754 TVAG 757
RESULT 10
ID ATSA_RAT
AC Q9ESP7; Q9ESP8; Q9ESP6; STANDARD; PRT; 630 AA.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ADAMTS-4 precursor (P3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
DE (Fragment).
GN ADAMTS4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Brain;
 RX MEDLINE=20415831; PubMed=10961658;
 RA Satoh K., Suzuki N., Yokota H.;
 RT "ADAMTS-4 (a disintegrin and metalloproteinase with thrombospondin motifs) is transcriptionally induced in beta-amyloid treated rat astrocytes.";
 RL Neurosci. Lett. 289:177-180(2000).
 CC -1- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE DESTRUCTION OF AGGREGAN IN ARTHRITIC DISEASES.
 CC -1- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-I-Ala-393 site.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).
 CC -1- TISSUE SPECIFICITY: ONLY EXPRESSED IN BRAIN.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AB042272; BAB16474.1; -
 DR EMBL; AB042271; BAB16473.1; -
 DR EMBL; AB042273; BAB16475.1; -
 DR InterPro: IPR001762; Disintegrin.
 DR InterPro: IPR001590; Reprolysin.
 DR InterPro: IPR000884; TSPL.
 DR InterPro: IPR000130; Zn_MTPeptidse.
 DR Pfam; PF00090; tsp.1; 2.
 DR SMART; SM00209; Reprolysin; 1.
 DR PROSITE; PS0215; ADAM_MEPRO; 1.
 DR PROSITE; PS0092; TSPL; 1.
 DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 DR Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
 KW Extracellular matrix.
 FT NON_TER 1 5
 FT PROPEP <1 5 BY SIMILARITY.
 FT CHAIN 154 630 ADAMTS-4.
 FT METAL 154 154 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 155 155 BY SIMILARITY.
 FT METAL 158 158 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 164 164 ZINC (CATALYTIC) (BY SIMILARITY).
 FT DOMAIN 233 303 DISINTEGRIN-LIKE.
 FT DOMAIN 316 367 TSP TYPE-1.
 FT DOMAIN 368 478 CYS-RICH.
 FT DOMAIN 479 630 SPACER.
 FT DOMAIN 40 45 POLY-ALA.
 FT CARBOHYD 96 96 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 474 474 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 630 AA; 68384 MW; 63A428753167CF CRC64;
 Query Match 40.6%; Score 1518.5; DB 1; Length 630;
 Best Local Similarity 55.8%; Pred. No. 1.4e-105;
 Matches 271; Conservative 67; Mismatches 137; Indels 11; Gaps 6;
 QY 209 RAKRFVSIPIYVETLVVADESQVYFGADLEHYLLTLLTAARLYRHPHSILLNPVIVVK 268
 DB 2 RTRKFASLSRFVETLVVADDKMAAFHAGLKHVLLTVMMAAAKAFKHPHSIRNPVNLVWVR 61

QY 269 VLLLRDRSGPKVTGNAAITLRFNFCAMQKLNKYSKHEPYWDFTAILFTRODLGATCTD 328
 DB 62 LVILSGQEVPOVGPVSAOITRSFCTWQGLNPNPNDSPDHFDTAILFTRODLGCVSTCD 121
 QY 329 TLGMADVGTMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKVEVFGK-LRAN 387
 DB 122 ALGMAGVGTVCDPARSCAIVEDDGLQSAFTAHELGHVFNHMLHNSKPCANLNGQGSRR 181
 QY 388 HMSPTLTIQIDRANPWSACSAITITDLDSDGHGDCILDQPSKPISLPEDLPFGASYTILSQ 447
 DB 182 HVMAPVMAHVDPEEPWSPCSARFITDFLDNGYGHCLLDKPEALHLPVTFPKDYDADQ 241
 QY 448 CELAFGVSGKPCPYM-QYCTKLMTGKAKQMVQCTRHFPWADGTSGEGLKLUKACVE 506
 DB 242 QLTFTGPDSSHCPQLPPPCAAALWCFHLNGHAMQCTHSPWADGTPCGPAQACMGRCCLH 301
 QY 507 RHNLNKHVD--GSWAKWDVPGSCRCGGVQGLARQCTNPTPANGKYGCEGVRYKRS 564
 DB 302 VDQLKDFNIPQAGMGPMGDCSRTCGGVQFSRRDCTKPVPRNGKYGCEGRRTFERS 361
 QY 565 CNLEPCSSASGKSFREPCQCEAFNGYNHSTNRLTL---AVAWPKYSGVSPDRCKLICR 621
 DB 362 CNTKNCP-HGSALTFREPCAA---YHRTDLEKSEFGPMQDWPRYTGTGAPRDQCKLTQ 417
 QY 622 ANGCTGYFVLAPKVVDTGLCSFDSVSCVQCKCIKAGCDGNLGSKKRFDKCGVCGDNKS 681
 DB 418 ARALGYVYLVPRVADGTPCSPDSSVSCVQGRCHAGCDRIIGSKKFKDKMCGGNGSS 477
 QY 682 CKKVTG 687
 DB 478 CSKQSG 483
 RESULT 11
 AT57_HUMAN
 ID AT57_HUMAN STANDARD; PRT; 997 AA.
 AC O9UKP4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7).
 GN ADAMTS7.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99395124; PubMed=10464288;
 RA Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
 RT "ADAM-TS5, ADAM-TS6, and ADAM-TS7, Novel Members of a New Family of Zinc Metalloproteases.";
 RL J. Biol. Chem. 274:25555-25563(1999).
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).


```

or send an email to license@isb-sib.ch).
-----
CC EMBL; AF140675; AAD56358.1; -
DR HSSP; P15167; 1ATL.
DR MEROPS; M12.231; -.
DR MIM; 605009; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Repolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF00090; tsp_1; 1.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPl; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00215; ADAM_MEPHO; 1.
DR PROSITE; PS00092; TSPl; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 232 BY SIMILARITY.
FT CHAIN 233 997 ADAMTS-7.
FT SITE 204 204 CYSTEINE SWITCH (POTENTIAL).
FT METAL 388 388 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 389 389 BY SIMILARITY.
FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 398 398 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 462 537 DISINTEGRIN-LIKE.
FT DOMAIN 538 594 TSP TYPE-1 1.
FT DOMAIN 595 697 CYS-RICH.
FT DOMAIN 698 914 SPACER.
FT DOMAIN 915 990 TSP TYPE-1 2.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 778 778 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED02FC104 CRC64;

Query Match 29.08; Score 1087; DB 1; Length 997;
Best Local Similarity 38.68; Pred. No. 3.2e-73;
Matches 286; Conservative 82; Mismatches 264; Indels 108; Gaps 32;

QY 1 MLLGLITLAFAGTAGGFEPEP---EWWPIRLDP-----DINGRR- 39
DB 18 LLLLCALAPGAPGAPGATEGRAALDIVHVRVDAGGSFLSYELWPAALKRDRVSRD 77
QY 40 ---YY---WRGPESDGLIFQITAFQEDFYHLTPDAQFLAPAF--STHGLGVPLQL 91
DB 78 APAFYELQYRGRE-----LRFNTAQ-----HL---LAPGFVSETRRRG---GL 116
QY 92 TGGSSDLR---RCFYSGDVNAEPD---SFAAVSLCGLRGAFGYRGAEVISPLN 144
DB 117 --GRAHIAHTPACHILGEVQ--DPELEGLLAISACDGLKGVFOLSNEFIPELD--SA 171
QY 145 PAAARNSSOAHLLQRRGV-----GGPSGDPTRSCGVAGSNPAIRALDYPKPRAGFG 199
DB 172 PAREPQAQ--PHVYKQQAERLAPLQAGDSSAP--STCGV-----QVYPELESRRERW 219
QY 200 ESRRRRSGRKRF-----VSPRVEVLVVADESQWVFHG--ADLEHYLLILLATAARLYR 254
DB 220 EQQRQRPRURLUHQRSVSKERKWTETLVADAKMYHGPQVQVESYVLTIMNVAGLFH 279
QY 255 HPSILNFINIVVKKVLLLRDRSGPKVTGNAALTLRNFCAWQKLNKVSQKHPEYWDTAI 314
DB 280 DPSIGNPHTITIVRLVLEDEEDLKITHADNTLAKSFCWKQKSNKMGDAHPLHDDTAI 339
QY 315 LFTQDLCGATT--CDTLGMADVTGMDPKRSCSVIEDDGLPSAFTTAHGLHVFNMPHD 372
DB 340 LLTRKDLCAAMNRPCTETLGLSHVAGMCQPHRSCSINEDTGLPLAFTVAHELGHLSFGIQHD 399
QY 373 -NVKVEEVFGKLANHHMSPFTLIQIDRANPWSACSNAIITDFDLSHGDCLLDQFSKP- 430

```

```

Db 400 GSGNDCBPV-GK--RPFIMSPQLLYDAAPTWSRCSQYITRFLDRGWGLCLDDPPAKDI 456
QY 431 ISLPEDLPGASYITLSQOCELAFGVGSKPCPYM-QYCTKLMCTGKAKGQWVCQTRHFPWAD 489
DB 457 IDPFSVPPGVLYDVSHQCRLOYGAYSACFEDMDNVCHTLWCSVGT----TCHSKLDAVD 512
QY 490 GTSCGEGKCLCKACVERHNLNKHVRDGSNAKWDPYGPCSRTCGGVQVLARRQCTNPTPA 549
DB 513 GTRCGENKWLCSGECVP-VGFRPEAVDVGWSGWSANSICSRSCGMGVQSAERQCTOPTPK 571
QY 550 NGKRYCEGVKVRYSNLEPCBSSASCKSPREEQCEAFNGYNHSTNRLFLAVAVWPKYSG 609
DB 572 YGRYCVGERKRFRLCNLQACP--AGRPSFRHVQCSHFDAMLYKGLQHLT-----WVPVND 625
QY 610 VSPDKCKLICRANGTGYFYVLAPKVVVDGTLCSPP--DSTSVQVQGVKICAKGDCNGLGSKK 667
DB 626 VNP---CELHCRPANEFYFAKKLRDAVVDGTPCYQVRAASRLDLCINGICKNVGCDFEIDSGA 682
QY 668 RFDKCGVCGGDNKSCKKVTG 687
DB 683 MEDRCGVCHGNGSTCHTVSG 702

RESULT 12
AT10_HUMAN
ID AT10_HUMAN STANDARD; PRT; 1077 AA.
AC Q9H324;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 10) (ADAM-TS 10) (ADAM-TS10) (Fragment).
GN ADAMTS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Apte S.S.;
RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
RT thrombospondin type I repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF163762; AAG35563.1; -
DR MEROPS; M12.235; -.
DR Genew; HGNC:13201; ADAMTS10.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Repolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF00090; tsp_1; 5.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR

```

DR SMART: SM00209; TSPI: 5;
DR PROSITE: PS00215; ADAM_MERO; 1.
DR PROSITE: PS00922; TSPI: 2.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT NON_TER 1 1
FT PROPEP 1 207 BY SIMILARITY.
FT CHAIN 208 1077 ADAMTS-10.
FT METAL 366 366 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 367 367 BY SIMILARITY.
FT METAL 370 370 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 376 376 ZINC (CATALYTIC) (BY SIMILARITY).
FT DOMAIN 434 520 DISINTEGRIN-LIKE.
FT DOMAIN 578 679 CYS-RICH.
FT DOMAIN 680 802 SPACER.
FT DOMAIN 521 577 TSP TYPE-1 1.
FT DOMAIN 799 860 TSP TYPE-1 2.
FT DOMAIN 862 918 TSP TYPE-1 3.
FT DOMAIN 922 976 TSP TYPE-1 4.
FT DOMAIN 981 1031 TSP TYPE-1 5.
FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 714 714 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 769 769 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 866 866 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1077 AA; 118072 MW; 3914DE18DCBBF587 CRC64;

Query Match 28.68; Score 1072.5; DB 1; Length 1077;
Best Local Similarity 36.88; Pred. No. 4.4e-72;
Matches 260; Conservative 89; Mismatches 280; Indels 77; Gaps 23;

QY 24 EVVPIRLDPDINGR-----RYWGPEDSGDGLIFQITAFQEDFYHLHTLDAQF 74
DB 14 EIAFPTVDH--NGALLAFSPPPRRRGRTGATAESRLFYKVASPTHELLNLTSSRL 71

QY 75 LAPARETEHLGVLPGTGGSSDLRCFYSGDVNAE--PDSFAVSLCGGLRGAFYRGAE 133
DB 72 LAGHVSVEYH--TREGLAWQARPHCLYAGHLOGQSSSHVAISTCGGLHGLIVADEEE 129

QY 134 YVISPPLNAPAAQNRNQQAHLLQRRGVGPGSGDPTSRGV--ASGW--NPAIRALRD 189
DB 130 YLIEPLHGGPKGRSPESGPHVYVYKSSLRHPLD--TAGGVDRDKPWKGRPWMLTLK 187

QY 190 PKPFRAGFGSRRSRGRKRVFIPRYVEITLVVADESMVKEHG--ADLEHYLLFLAT 248
DB 188 P--PPARPLGNETERGQPG--LKRSVSRERYETLVADKMWAYHGRDRVQYVLAIMNI 244

QY 249 AARLYRHPISILNPIVIVVVKLLLRDSDGPKVTGNAALTLRNFCAWKKL-----NK 301
DB 245 VAKLFQDSLSGSTVNLIVRLITLITEOQPLEITHHAGKSLDSECKWQKSIYVHSHGNA 304

QY 302 VSDKHPEYWDYAILFTRODLC--GATCTDILGMADVTCMDPKRSCSVIEDDGLPSAFT 359
DB 305 IPENVANHTAVLTIDYICIKYKPCGTGLAPVGMGERSCSVNEDIGLPOAFTI 364

QY 360 AHELGHVFNMPHDNVKVEEVEFG---KLRANHMKSTPLIQRANP--WSACSAALITD 413
DB 365 AHEIGHTFGMHNDGVNCGARGQDPKALMAAHTMTK-----NPFVWSSCNRYDITS 417

QY 414 FLDSGHGCLLDQPSK--PISPELPGASYTLSSQCELAFAFGVSKPCPYMQYCTKLWCTG 472
DB 418 FLDSGLGLCLNRRPPQDEVYPTVAPQAYDADQCFQHGKVRQSKYGEVCSSELWCLS 477

QY 473 KAKQWVQCTHFHWADTSCGEKL-----CLKGACVVERINLKHVRDGSNAKWDYGP 528
DB 478 KSNR---CITNISPAEETLCQTHIDKGMCYKRVCP--EGSRPEGVDGAWGFWTPWGC 533

QY 529 SRTCGGGVQLARRQCTNPTPANGKYGEGVVRKVRSCNLEPCPSSASGKGFREOQCEAFN 588
DB 534 SRTCGGGVSSSSRRHCDSPRTIGKCYCLGERRRHRSCTDDCPGS--QDFREVQCSFED 591

QY 589 G-----YNHSTNRLRLAVAWVPKYSGVSPROKCKKLCIRANGTGYFYVLAPKVDGTL 641
DB 592 SIPFRGKYKWKTYR-----GGGVK---ACSLTSLAEGNFYTERAAAVVDGTPC 638
QY 642 SPDSTSVGVQCKICAKAGDNLGSKKRFDKGCGGDNKSKKVTG 687
DB 639 RPDVTVCVSGECKHVGCDRLVLSGLREDKRCVCGGDSACETIEG 684

RESULT 13
AT12_HUMAN
ID AT12_HUMAN STANDARD; PRT; 1593 AA.
AC P58397;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=21264577; PubMed=11279086;
RA Cal S. Argueelles J M, Fernandez P L, Lopez-Otin C.;
RT "Identification, characterization, and intracellular processing of ADAM-TS12, a novel human disintegrin with a complex structural organization involving multiple thrombospondin-1 repeats.";
RL J. Biol. Chem. 276:17932-17940(2001).
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (by similarity).
CC -!- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely expressed in gastric carcinomas and in cancer cells of diverse origin.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -!- PTM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1 DOMAINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL; AJ250725; CAC20419.1; -
Genew; HGNC:14605; ADAMTS12.
MIM: 606184; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR000130; Zn_MTPetdse.
DR Pfam; PF00090; tsp_1; 6.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPI; 8.

Lapierre C.M.;
"cDNA cloning and expression of bovine procollagen I N-proteinase: a new member of the superfamily of zinc-metalloproteinases with binding sites for cells and other matrix components."; *Proc. Natl. Acad. Sci. U.S.A.* 94:2374-2379(1997).
[2]
RP PARTIAL SEQUENCE.
RX MEDLINE-95348096; PubMed-7622483;
RA Collige A., Beschlin A., Samyn B., Goebels Y., Van Beeumen J.,
Nusgens B.V., Lapierre C.M.;
RT "Characterization and partial amino acid sequencing of a 107-kDa
procollagen I N-proteinase purified by affinity chromatography on
immobilized type XIV collagen."; *J. Biol. Chem.* 270:16724-16730(1995).
RL J. Biol. Chem. 270:16724-16730(1995).
CC -I- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
CC COLLAGEN BIOSYNTHESIS.
CC -I- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
CC alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains
CC at Ala-1-Gln.
CC -I- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -I- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
CC COLLAGEN TYPE XIV.
CC -I- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -I- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVEL
CC IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS
CC AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS
CC WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.
CC -I- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -I- PTM: THE N-TERMINUS IS BLOCKED.
CC -I- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -I- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF DERMATOSPARAXIS, A
CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN
CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN
CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -I- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -I- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -I- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
CC or send an email to license@isb-sib.ch).

DR EMBL; X96389; CAA65253.1; -
DR MEROPS; M12.301; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_Mtpeptdse.
DR Pfam; PF00090; tsp_1; 4.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSP1; 4.
DR PROSITE; PS500215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSP1; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 28 POTENTIAL.
FT PROPEP 29 253 BY SIMILARITY.
FT CHAIN 254 1205 ADAMTS-2.
FT METAL 402 402 ZINC (CATALYTIC) (POTENTIAL).

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 41.8928 Seconds
(without alignments)
3378.970 Million cell updates/sec

Title: US-10-009-332-1-copy_1_687
Perfect score: 3744
Sequence: 1 MLLGLITLAFAGTAGGFE.....RFDKCGVCGGDNCKKVTG 687

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mmc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3736	99.8	950	4 Q8TE58	Q8te58 homo sapien
2	1223	32.7	2165	5 Q19791	Q19791 caenorhabdi
3	1022	27.3	1688	5 Q8SX80	Q8sxb0 drosophila
4	1015	27.1	1081	4 Q8TE60	Q8te60 homo sapien
5	1015	27.1	1034	5 Q9W493	Q9w493 drosophila
6	998.5	26.7	1095	4 Q8TE56	Q8te56 homo sapien
7	989.5	26.4	1229	5 Q9VF61	Q9vf61 drosophila
8	967.5	25.8	1072	4 Q8TE57	Q8te57 homo sapien
9	941.5	25.1	1207	4 Q8TE59	Q8te59 homo sapien
10	928	24.8	1223	4 Q8WXS8	Q8wxs8 homo sapien
11	928	24.8	1223	4 Q8TE55	Q8te55 homo sapien
12	926.5	24.7	1159	4 Q8TEY8	Q8tey8 homo sapien
13	830	22.2	269	6 Q9GL54	Q9gl54 oryctolagus
14	720	19.2	1427	4 Q96L37	Q96l37 homo sapien
15	558	14.9	192	6 Q95N24	Q95n24 equus cabal
16	518	13.8	187	6 Q95N23	Q95n23 equus cabal

17	501.5	13.4	1062	5 Q19204	Q19204 caenorhabdi
18	497.5	13.3	1091	5 Q9W126	Q9w126 drosophila
19	479	12.8	1444	5 Q17591	Q17591 caenorhabdi
20	426.5	11.4	872	5 Q22580	Q22580 caenorhabdi
21	411	11.0	790	5 Q8T458	Q8t458 drosophila
22	406.5	10.9	2167	5 Q76840	Q76840 caenorhabdi
23	400	10.7	340	11 Q91256	Q91256 mus musculu
24	389	10.4	364	4 Q9UGQ1	Q9ugq1 homo sapien
25	383	10.2	1572	5 Q4938	Q4938 haemochus
26	374	10.0	766	4 P82987	P82987 homo sapien
27	373.5	10.0	899	13 Q8UVF1	Q8uvf1 coturnix co
28	358	9.6	525	4 Q96RW4	Q96rw4 homo sapien
29	356	9.5	1235	4 Q95428	Q95428 homo sapien
30	353	9.4	1280	11 Q9EPX2	Q9epx2 mus musculu
31	352	9.4	610	13 Q93523	Q93523 bothrops ja
32	352	9.4	610	13 Q8QG88	Q8qg88 bothrops in
33	339	9.1	610	13 Q9Y120	Q9y120 agkistrodon
34	339	9.1	1014	5 Q95R33	Q95r33 drosophila
35	338	9.0	192	11 Q9CX59	Q9cx59 mus musculu
36	337.5	9.0	609	13 Q9W6M5	Q9w6m5 agkistrodon
37	337	9.0	609	13 Q90282	Q90282 croatalus at
38	332.5	8.9	105	6 Q9GLK7	Q9glk7 oryctolagus
39	332.5	8.9	479	13 Q9PWJ0	Q9pwj0 agkistrodon
40	331.5	8.9	922	13 Q8UVF2	Q8uvf2 coturnix co
41	326	8.7	466	13 Q91AX7	Q91ax7 agkistrodon
42	323.5	8.6	617	13 Q90499	Q90499 echis pyram
43	322.5	8.6	3198	5 Q908G8	Q908g8 manduca sex
44	322	8.6	477	13 Q98SP2	Q98sp2 bothrops ja
45	322	8.6	620	13 Q42138	Q42138 agkistrodon

ALIGNMENTS

RESULT 1

Q8TE58	PRELIMINARY;	PRT;	950 AA.
ID	Q8TE58		
AC	Q8TE58;		
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Metalloprotease disintegrin 15 with thrombospondin domains.		
GN	ADAMTS15.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21856482; PubMed=11867212;		
RA	Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,		
RA	Lopez-Otin C.;		
RT	"Cloning, expression analysis, and structural characterization of		
RT	seven novel human ADAMTSs, a family of metalloproteinases with		
RT	disintegrin and thrombospondin-1 domains.";		
RL	Gene 283:49-62(2002).		
DR	EMBL: AJ315733; CAC86014.1; -		
DR	Integrin; Protease.		
SK	SEQUENCE 950 AA; 103286 MW; 5DFBE18285CCCC3B CRC64;		
QY	Query Match 99.8%; Score 3736; DB 4; Length 950;		
QY	Best Local Similarity 99.9%; Pred. No. 1.5e-302;		
QY	Matches 686; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	1 MLLGLITLAFAGTAGGFEPEPEVVPVIRLDPDINGRRYWRGPDSDGQLFIQTAF 60		
Db			
Db	1 MLLGLITLAFAGTAGGSEPEPEVVPVIRLDPDINGRRYWRGPDSDGQLFIQTAF 60		
QY	61 QEDFYHLTPDAQFLAPAFSTEHLPVLPGLTGGSSDLRRRCFYGDVNAEPDPSFAVSLC 120		
Db			
Db	61 QEDFYHLTPDAQFLAPAFSTEHLPVLPGLTGGSSDLRRRCFYGDVNAEPDPSFAVSLC 120		
QY	121 GGLRGAFGYRGAEYVISPLPNASAPAAQNSQGAHLQRRGVPGSPGDPTRSGCVASGW 180		

```
|||||
Db 121 GGLRGAFYGAAYVISPNSAPAAQNSOGAHLQRRGVPGPSGDPTSCRCVASSW 180
QY 181 NPAILRALDPYKPRRAGFESRRSRGRKRFVSPRYVETLVVADESVMKFFHGADLEH 240
Db 181 NPAILRALDPYKPRRAGFESRRSRGRKRFVSPRYVETLVVADESVMKFFHGADLEH 240
QY 241 YLLTLATAARLYRHPSILNPINIVVVKLLLRDRSDGPKVTGNAALTIRNFCAMQKLN 300
Db 241 YLLTLATAARLYRHPSILNPINIVVVKLLLRDRSDGPKVTGNAALTIRNFCAMQKLN 300
QY 301 KVSQKHPEYWDTAITRQDLGCAITCDTLGMADVTGMDCKRSCSVIEDDGLPSAFTTA 360
Db 301 KVSQKHPEYWDTAITRQDLGCAITCDTLGMADVTGMDCKRSCSVIEDDGLPSAFTTA 360
QY 361 HELGHVFNPHDNNVKEEVFGKLRANHMSPITLQIDRANPWSACSAIITDFLDSHG 420
Db 361 HELGHVFNPHDNNVKEEVFGKLRANHMSPITLQIDRANPWSACSAIITDFLDSHG 420
QY 421 DCLLDQSPKISLPEDLPASVYLSQOCELAFGVGSKPCPYMQYCTKLVCTGKAKQWVC 480
Db 421 DCLLDQSPKISLPEDLPASVYLSQOCELAFGVGSKPCPYMQYCTKLVCTGKAKQWVC 480
QY 481 QTRHFPWADGTSQGEKCLKLGACVERHNLNKHVRDGSWAKWDPYPCSRCTCGGVQLAR 540
Db 481 QTRHFPWADGTSQGEKCLKLGACVERHNLNKHVRDGSWAKWDPYPCSRCTCGGVQLAR 540
QY 541 RQCTNPTPANGKYCEGVVRKYRSCNLEPCPSASGSKSFREOCEAFNGYNHSTNRLTLA 600
Db 541 RQCTNPTPANGKYCEGVVRKYRSCNLEPCPSASGSKSFREOCEAFNGYNHSTNRLTLA 600
QY 601 VAWPKYSGVSPDRCKKLCIRANGTYFYVYLAAPKVVVDGTLCSPDSTSVCVQKCIKAGCD 660
Db 601 VAWPKYSGVSPDRCKKLCIRANGTYFYVYLAAPKVVVDGTLCSPDSTSVCVQKCIKAGCD 660
QY 661 GNLSKRRFDKCGVCGGDNKSKCKVTG 687
Db 661 GNLSKRRFDKCGVCGGDNKSKCKVTG 687

RESULT 2
Q19791 PRELIMINARY; PRT; 2165 AA.
AC Q19791; Q27524;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE F25H8.3 protein.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Gajadaty S.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock J., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
RL Nature 368:32-38(1994).
```

```
EMBL; Z69361; CAA93288.1; -.
EMBL; Z69360; CAA93288.1; JOINED.
EMBL; Z69360; CAA93287.1; -.
EMBL; Z69361; CAA93287.1; JOINED.
DR HSP; P15167; IDTH.
DR MEROPS; M12.135; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR InterPro; IPR00130; Zn_Mtpeptidse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00080; tsd_1; 14.
DR SMART; SM00209; TSPI; 18.
DR PROSITE; PS00215; ADAM_MEPPO; 1.
DR PROSITE; PS00092; TSPI; 6.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 2165 AA; 244397 MW; FCC3DA8AAA9C4888 CRC64;

Query Match 32.7%; Score 1223; DB 5; Length 2165;
Best Local Similarity 40.5%; Pred. No. 1.4e-92;
Matches 257; Conservative 87; Mismatches 214; Indels 76; Gaps 16;

QY 101 CFYSGDVNAEPDSFAAASLC---GGLRGAFYGAAYVISPNSASA---PAAQRNSOGA 154
Db 159 CIYFAHVKG-VHSHIVNLCDSDEGLYGLMALPSGIHTVEPIISGNGTEHDGASRHRQ-- 225
QY 155 HLLQRRGVPGPSGDPTSCRGV-----ASGNPAILLRALDPYKPRRAGFESRSR 204
Db 226 HLVRKFDPMHFKSFDFHLSNSTSVNETTETVATWQDQWEDVIER-----KARSR 272
QY 205 RRSGRAKRFVSPRYVETLVVADESVMKFFHGADLEHYLLTLATAARLYRHPSILNPINI 264
Db 273 RAAN-----SWDHVVEVLVVDATKMYEYHGRSLEDYVLTLSFVASTYRHSRASINV 326
QY 265 VVVKYLLLRDRSDGPKVTGNAALTIRNFCAMQKLNKVKSDKHPEYWDTAITRQDLGCA 324
Db 327 VVVKLVLTENAGPRITQNAQOTLQDFCRWQYYNDPDDSSVQHHDAVAILLTRKDCRS 386
QY 325 T-TCTLTGMADVTGMDCKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNNVKEEV--- 380
Db 387 QGKCDTLGLAELGTMCMDQKSCAIIEDNGLSAAFTIAHELGHVFSIPHDDERKCTYMPV 446
QY 381 -----FGKLRAN---HMSPTLIQIDRANPWSACSAIITDFLDSHG--DCLLD 425
Db 447 NKVKCFQSTKFKDKTOFQNNFHIMAPTLEVNTHPWSWSPSCSAGMLERFLENNRGOTQCLFD 506
QY 426 QPSKPSILPE-----DLPGASYTLSSQOCELAFGVGSKPCPYMQYCTKLVCTGKAKQWVCQ 481
Db 507 QPVERRYVEDVFRDEPKKYYDAHQCKFVFGPASELCFPMPTCRLRLWCATFYGSQMGR 566
QY 482 TRHFPWADGTSQGEK--LCLKGACVERHNLNKHVRDGSWAKWDPYPCSRCTCGGVQLA 539
Db 567 TQHPWADGTPCDESRSMFCHHGACVRLAPESLTAKIDGQWDMRSWGECRSRTCGGVQKG 626
QY 540 RRQCTNPTPANGKYCEGVVRKYRSCNLEPCPSASGSKSFREOCEAFNGYN-----HS 593
Db 627 LRDCDSPKPRNGKVCVQORERYRSCNTQECPWDT--OPREVQCSSEFNKNDIGIQGVAS 684
QY 594 TNRLTLAVAWVPKYSGVSPDRCKKLCIRANGTYFYVYLAAPKVVVDGTLCSPDSTSVCVQK 653
Db 685 TN-----THWPKYANVAPNERCKLYCRUGSAAFTLLRDKVVDGTPCDRNGDDICVAGA 739
QY 654 CIKAGDGNLGSKKRFDKCGVCGGDNKSKCKVTG 687
Db 740 CWPAGCDHQLHSTLRDKCGVCGGDDSSCKVKVG 773

RESULT 3
Q8SXB0 PRELIMINARY; PRT; 1688 AA.
ID Q8SXB0
AC Q8SXB0;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
```



```
DE GH16393p.
GN CG6107.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Paciel J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AV094716; AAM11069.1;
SQ SEQUENCE 1688 AA; 189867 MW; 48FB8DD4D50C4AD2 CRC64;

Query Match 27.3%; Score 1022; DB 5; Length 1688;
Best Local Similarity 35.1%; Pred. No. 5.8e-76;
Matches 237; Conservative 104; Mismatches 237; Indels 98; Gaps 24;

QY 55 FOITAFQEDFYHLTPDAQFLAPAFSTEHLGVPLQG-----LTGGSSDLRR--CFYSGD 106
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db YNLNFGRLHLVLRQDASFV--HNHSMTHIRILKEGHEHPGPEAEAEQRHLGCFYSGY 354
QY 107 VNAEPDSFAAVSLCGLRGAFYGAEVYISPLPNASAP-----AAQNSOGAHLQ 158
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db VEDDPHSMVSVSLCGWGTGYKTSFGALLIOPVNTSDEVLHVRFSQRNAR--HAYS 412
QY 159 RRGVPGGSDPTSCGVASGNWPAIRALDPYKPRRAGFGE---SRSSRRSGRAKRFV 214
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db GLDDFMSKLVQVEEQKSKSRKLNKRRIYA 448
QY 215 STPRVY---ETLVVADESVKPHGADLEHYLLTLTAAARLYRHPSLNINIVVVKLL 271
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db DVDNQVYTLVLIADVNSMKOPFGEDLQPIYLILMSIVSSIFADASIGNSRILVRLIS 508
QY 272 LDRDRSGPKV---TGNAALTIRNFCAWQKLNKVSDEKHPYWDATLFTRODLGAT--- 325
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db L-----PNINDQTHSSNEMLKHFCF---INQSYER---DTAMLITREPICGSVPGK 555
QY 326 TCDTIGMADVGMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVVKVCEEVFGKLR 385
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db ICHMLGLAELGTVCV--SSSCSIVQDTGLPTAFTMAHELGHILNNHDDDDCKMPYVTRQN 614
QY 386 AN---HMSPTLIQIDRANPWSACSAAIITDFDSHGDCDCLLDQSPKISLP---EDLP 439
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db NKVLHIMSVMGIIHMPWMSKCSRHFVSEFLEKTDKSCV---ETSVGAHPIYGTRELPG 672
QY 440 ASYTLSSQCELAFGVSGKPCPYMOYCTKLWC---TGKAGQWVCOTRFPWADTSCG-E 495
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db EIYSLDAQCLSGNDFGCTPDECKRLWCNRTSGNSNEQ--CASSNLPWADGTGCGSS 730
QY 496 GKCLKGACVERHNLNKH-----RVDGSAKWDPYGPCSRTCGGVQVLARRQCTNPTFANG 551
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 GHWCQRGKCVS---NKHGYGRQVNGGWPWTPTPCSLTCGGGVQESRRRCNQPVPENG 786
QY 552 GRYCEGVVYKVRSCNLEPCPSASGKSPREOCEAFNGYNYHSTNRLTLAVAWVYKSGVS 611
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 787 GYCTGSRKKYRSCNTHQCPGSMDDP---REQOCYAMGRNNINIPGVNPTDKWVPKYE--- 841
QY 612 PRDKCLICRANGTGYFYVLAPKVVVDGTLCPSPDSTSVCOGKCIKAGDCGNLGSKKRFDK 671
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 842 -KDACLFCRMDKMYTFMLKSWYTDGTSCAVDSFDKCVNGICRPAGCDNENLSIAKLDK 900
QY 672 CVCVCGDNNKSKVTG 687
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 901 CVCVCGRNDTCHVTG 916
```

```
RESULT 4
Q8TE60 PRELIMINARY; PRT; 1081 AA.
AC Q8TE60;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE ADAMTS18 protein.
GN ADAMTS18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002)
DR EMBL: AJ311903; CAC83612.1;
SQ SEQUENCE 1081 AA; 119655 MW; 0438BF645676461E CRC64;

Query Match 27.1%; Score 1015.5; DB 4; Length 1081;
Best Local Similarity 33.9%; Pred. No. 1.1e-75;
Matches 255; Conservative 98; Mismatches 266; Indels 133; Gaps 30;

QY 16 AGGPEPEPREVVPIRLDP-----DI--NGRRYMRGPEDSGDGLIFQITAFQEDFY 66
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 54 ASGLNDDYVFTPEVDSAGSYIISHDLHNGRK---KRSQAQNASLSLHYRFSAGFQELHL 110
QY 67 HLPDPAQFLAPAFSTEHLGVPLQGLGSS---DLRCFYSGDVNAEPDPSFAVSLCG 122
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 ELKPS-----AILSSHFIQVGLKDGASETQKPEVQCFQYQFIRNDSSSSVAVSTCAG 164
QY 123 LRGAFGYRGAEYVISPULP-----NASAPAAQNSOGAHLQ-----RGVPG- 164
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 LSGLIRTKNEFLISPLPQLLAQEHNVSSPAGHH---PHVLYKRTAEKIQIRYGP 220
QY 165 -----GPSGDP---TSR-----CGVASGNWPAIRALDPYK----- 193
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 221 GRNPGYSPSHIPHASOSRETYHRRLOKHFCGRKKYAP-----KPPTEDTYL 271
QY 194 RRAGFGES--RSRRSGRAKRFVSIPTVETLVVADESVKPHG--ADLEHYLLTLTAAAR 251
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 272 RFEYSGSGRRPRRSAGSKQKGLN-----VETLVVADKKMKVEKHGKGNVTYILTVMKVS 326
QY 252 LYRHPSTLNPINIVVVKVLLLRDRDSDGPKVTGNAALTIRNFCAWQKKL--NKVSDKHPEY 310
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 LFXDGTIGSDINVVVYSLILLEQEPGGLLNHHADSLNSFCQWQSAALIGKNGKRH--- 382
QY 311 DTAITFRQDLG--ATTCDTLCMADVGMCDPKRSCSVIEDDGLPSAFTTAHELGHVFN 368
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 DHAITLTGFDICSNKNEPCDTLGFAPTSGMCKSYRSTINEDTGLGLAETIAHESGNFG 442
QY 369 MPHNV--KVCVEVFGKLRANHMSPTLIQIDRANPWSACSAAIITDFDSHGDCDCLLDQ 427
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 443 MIHDGEGNCPKAEK-----NIMSPITLGNNGVFSWSSCSQYLLKFLSTPQAGCLVDEP 497
QY 428 SK--PISLPEDLPASYYLSQCELAFGVSGKPCP---YMQYCTKLWC--TGKAGQWVC 480
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 498 KQAGQYKPKLPGQIYDADTQCKWQFQAKAKLCSLGFVKDICKSLWCHRVGHR-----C 552
QY 481 QTRHFWADTSCGEGKCLKGACVERHNLNKHVRDGVSNKAWDPYGPCSRTCGGVQVLAR 540
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 553 ETKFMAAEGTVCGLSMWCRCQGCQVFGELGPRPIHGQWSAWSKSECSRTCGGVKFGQE 612
QY 541 RQCTNPTPANGKYGCVRVKVRSCNLEPCPSSASGKSPREOCEAFN-----GYNHSTN 595
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 613 RHCNPKPYGGIFCPGSSRIYQLCNINPC--NENSLDFRAQOQCAEYNSKPFGRWFYQ-- 668
```

QY 596 RLTLAVAWPKYSGVSPRDKKLCIRANGTGYFYVLAPKVVDTGLCPDSTSVCVQKCI 655
Db 669 -----WKP-YTKVEEDRCKLYCKAENFEFFAKGVKVDGTPCSPKNDVCDIGVCE 720
QY 656 KAGCDGNLGSKRFRKCGVCGGDNKSKKVTG 687
Db 721 LVGCDHELGSKAVSDACGVCGKGDNSTCKFYKG 752

RESULT 5
Q9W493 PRELIMINARY; PRT: 1054 AA.
AC Q9W493;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE CG4096 protein.
GN CG4096.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila;
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers J.R., Blazef R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
Ra de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista A.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
RA Glodek A., Glendon A., Glendon J.H., Gu Z., Guan P., Harris K.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle B.J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner T.C., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003435; AAF46065.1; .
DR HSP; P15167; IATL.
DR MEROPS; M12.231; .
DR FlyBase; FBgn029791; CG4096.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Repolysin.
DR InterPro; IPR000884; TSPL.
DR InterPro; IPR000130; Zn_Mtpeptidase.

DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Repolysin; 1.
DR Pfam; PF00090; tsp_1; 2.
DR SMART; SM00209; TSPL; 2.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSPL; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN.1.
SQ SEQUENCE 1054 AA; 118616 MW; DC15455555CB6212 CRC64;

Query Match 27.1%; Score 1015; DB 5; Length 1054;
Best Local Similarity 33.2%; Pred. No. 1.2e-75;
Matches 266; Conservative 93; Mismatches 269; Indels 174; Gaps 29;

QY 17 GGFE-----EREVVPRLDPD-----INGRRYVWR-----GPED 47
Db 23 GGVRLYGLHSDLVAGEQLVPRVRHPDGAFTWOLEYAHLEDRHRQRSLNSEHD 82
QY 48 SGDOGLIFQITAFQEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSDRLRCFFYSGDV 107
Db 83 TQADLHLILLPLANETLHLELMAHSYFLAPNLVVERHRRDLRTSPUTTRHLNCHPHGV 142
QY 108 NAEPSFANVSLCGL-----RGAFYRGAEVVISPL-----PNASA 144
Db 143 RGQATNVAISTCAGLVSCPFCVLPILITLCQVGHIRTAGNEFYIEPSEHEHPVNGH 202
QY 145 P--AAQRNS--OGAHLQRRG--VFGPSGD--PTSRCGVASGMNPAILRALDPYKPR-- 195
Db 203 PHVFORSVKPKHSLRKRNRKRGKSGSGAEVSNCGTR-----EPRRM 248
QY 196 -----AGGES 201
Db 249 ETRLEWQARGKVQGGQIRRRHHHHHHKHYRHHOOKISRVPHTKFYEQFTE 308
QY 202 RSRRSRGAKEFVSIPRVETLVVADESMVFGHGADEHYLLTLATAARLYRPSILNP 261
Db 309 PDHAIEPRRRSSISPRVETLIVADATMSAFH--RDLNGYLLTLMNVSALYKDPDSGNS 367
QY 262 INIVVVKVLLLRDRDSGPK--VTGNAALTLRNFCAWQKLNKVDKPIEYWDTAILETRQ 319
Db 368 IEIVVRIIQLDEESQQLNLTQNAKNLDRFCSWQHLKNGSEKDPHHHDVAILLTRK 427
QY 320 DICGATTCDTLGMADVTCMPKSCSVIEDGLPSAFTHLGHVFMNPHDNVVKV-CE 378
Db 428 NTC--ANNCWTLGLANVGMCKPKQSCSVNEDNGIMLSHTITHLGHNFPMFHTAKIGCH 486
QY 379 EYFGKLRANHMSPL-----IQRANWSACSAAIITDFLDSGHGDCLLDQPSKPI--- 431
Db 487 PRVGPPI--VHIMTPTFGADTLQV-----CWSNCSRKYIITFLDQGLGEC--LDDPTPLDEY 539
QY 432 SLPELDLPGASYTLQQCELAFG-----VGSKPCPYMOYCTKLWCTKAKQMWVQCTRHF 485
Db 540 NTGELPGMYRNARGCRLQFNLTDTSEVGACSAFH-EFCSTLWC--KVNGE--CVTHMR 594
QY 486 PWADGTSCGEGKLCILKAGACVERHNLNKHVRVDSWAKWDPYPCSTTCGGVQLARRQCTN 545
Db 595 PTAPGTLGRNKWCNGKQVRREEL--AAVNGGWGDWSEWSECSRSCGGVSTQOQRECDN 652
QY 546 PTPANGKCYCEGVRYKYRSCNLEPCPSSASGSKSFREEOCEAFN--GYNHSTNRLTLAVAW 603
Db 653 PVPANGVFCIGERKRYKICRKRPCP--AEPSFRQAQCARFEDNVSQYATK-----W 704
QY 604 VPKYSGVSPRDKKLCIRANGTGYFYVLAPKVVDTGLCPDSTSVCVQKCIKAGCDGNL 663
Db 705 LPFFDKNPP--CKLFCSDVDDTIANWGATVLDGTPCTGLTNNNICIDGICKKVGCDWIV 761
QY 664 GSKKRFKCGVCGGDNKSKKY 685
Db 762 DSEVQDRRCVCGCGGSDGQCPV 783

RESULT 6
Q8TE56 PRELIMINARY; PRT: 1095 AA.
ID Q8TE56

```
AC Q8TE56;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Metalloprotease disintegrin 17, with thrombospondin domains.
GN ADAMTS17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ315735; CAC86016.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1095 AA; 121099 MW; A5C6540484754D5D CRC64;

Query Match 26.7%; Score 998.5; DB 4; Length 1095;
Best Local Similarity 34.3%; Pred. No. 2.9e-74;
Matches 253; Conservative 106; Mismatches 285; Indels 93; Gaps 26;

QY 2 LLLGILTLAF----AGRTAGGPEPEREVVPIRLDPDI-----NRRRYWRGPE 46
DB 10 LVPLVLLLVWGLDPTGVGDAADVEVLPWRVRPDDVHLPPLPAAPGPRRRRPTTP 69
QY 47 DS-----GDQGLIFQITAFQEDFYLLHLPDAQFAPAFSTHGLVPLQGLTGGSDLRRC 101
DB 70 AAPRARPGERALLHLPAFGRLYLQLRDLRFLSRGSEVEEAGARR--RGRPAEL--C 125
QY 102 FYSQDVNAEPDFAFVSLCGGLRGAFG--YGAEEV--ISPLPNASAPAAQNSOGAHLQ 158
DB 126 FYSGRVLGHGSLVLSACGAAGGLVGLIQGQEVQLPLNNSQGFSGRE----HLIR 181
QY 159 RR-GVPGPGSGD---PTSRCGVASGWNPAILRALDPYKPRAGFCE--SRSRRRSGRAKRF 213
DB 182 RWSLTPSPSAEAPQPEQLCKVLT-----EKKKPTWGRSDRWEERNAILR 228
QY 214 VSIPIRYVELLVADSMVKFGAD--LEHYLTLTATARLVHRPILNPINIVVKVLL 272
DB 229 TS-EHTVTLVADADVMYQYHGAEAARFILTVMNVMYMFQHSGLGKINIQVTKLVL 287
QY 273 RDRDSGPVKTGNAALTNRNFCWQKK-----LNKV--SDKHPEWDTAILFTRODL 321
DB 288 RQPAKLSIGHGERSLESFCHWQNEEYGGARYLGNQVPGCKDDPLVLDRAVFTTDF 347
QY 322 C--GATTCDTLGMADVGTMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPMHNVKVEE 379
DB 348 CVHKDEPCTVGIAYLGVCSAKRCKVLAEDNGLNLAFTIAHELGNLGNMHD----DD 402
QY 380 VFGKLRAHNSPTLIQIDRANP---WSACSAIIITDFLDSGHGDCLL----DQSPKPS 432
DB 403 HSSACGRHISMGEWVK--GRNPDSLSSWSSCRDLENFLSKSVSTCLVLTDRPSQTVR 460
QY 433 LPEDLPGLASYTLISQCEALFAGVGSKPCPYMAY--CTKLWCTGKAKGMVQCTREHPWADG 490
DB 461 LPHKLPGMHYSANEQCQILFGMNATFCRNMEHLKMCAGLWCL--VEGDTCKTKLDPLDGL 518
QY 491 TSCGEGKLCILGACVVERNLNKHVRDGSWAKWDYPGCPSRFCSTGGGVQLARRQCTNPTPAN 550
DB 519 TECGADKWCRAKECVSKTPIPEH--VDGDSWPWGAWSMCSRTCGTGARFRKCDNPPPGP 577
QY 551 GKGYEGRVYKVRNCLNPPSSASGSKSFREEQCEAFNGYHNSHNRILTAVAVPKYSGV 610
DB 578 GGTHCPGASVHAVCNELPCPKGL--PSFRDQQAQHDRUSPKKGLLTAV-----V 627
QY 611 SPRDKCKLICRANGTYGYVILAPKVVVDGTLCSPDSTSVCGQKCIKAGCDNGLGSKKRF 670
DB 670
```

```
DB 628 VDDPCPELYCSPLGKESPLLVADRVLDGTGCGPYETDLCVHGKCKQKIGDGIISAAKED 687
QY 671 KCGVCGDGNKSKKVTG 687
DB 688 RCGVCSGDGKTCHLVKG 704

RESULT 7
Q9VF61
ID Q9VF61 PRELIMINARY; PRT; 1229 AA.
AC Q9VF61;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE CG6107 protein.
GN CG6107.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J., Andrews-Pfannkoch C., Baldwin D.,
RA Arif J.F., Agbayani A., An H.-J., Ainsworth P., Brotherton P.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.I., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003709; AAF55199.1; -.
DR FlyBase; FBgn0038340; CG6107.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPL.
DR InterPro; IPR000130; zn_MTpeptdse.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp.1; 2.
DR SMART; SM00209; TSPL; 3.
```

DR PROSITE; PS500215; ADAM_MEPRO; 2.
DR PROSITE; PS50092; TSP1; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Zinc.
SQ SEQUENCE 1229 AA; 138136 MW; C1293180AB19B700 CRC64;

Query Match 26.4%; Score 999.5; DB 5; Length 1229;
Best Local Similarity 35.0%; Pred. No. 1.9e-73;
Matches 237; Conservative 102; Mismatches 227; Indels 111; Gaps 28;

QY 55 FQITAFQEDFYHLPTDAQFLAPAFSTEHLGVPLQG-----LTGGSDLLR--CFYSGD 106
DB 195 YNLNVEGRLHVLRLDASV--HNHSMTHIRILKGEHGPETAEAEQRHLGCFYSGY 253
QY 107 VNAEPDFAAVSLCGGLRGAFYGAEEYISPLPNASAP-----AAQRNSQGAHLQ 158
DB 254 VEDDPHSMVSVSLCGGMTGYIKTSFCALLIQPNVTSDEVLHRYFRKSQARNAR--HAVS 311
QY 159 RRGVPGGSDPTSCGVASGWNPAILRALDPYKPRAGFGE-----SRSRSSGRAKRFV 214
DB 312 KPFL-----GLDDFMKLEQVQEEQKSKRLNRKRHYA 347
QY 215 SIPRYV---ETLVVADESMVKFEGADLEHYLLTLLATAARLYRHPISILNPIVIVVKLL 271
DB 348 DVDNQYTLLEVLIADVNSMKQFHGEDLQPIILILMSIVSIFADASIGNSIRILLVRLIS 407
QY 272 LRDRSGPKV---TGNAALTLRNFCWQKLNKVDKHPYWDYDTAILFTQDLGAT--- 325
DB 408 L-----PNINDQTHSSNEMLKHFQF---INQSYER---DTAMLTIREPICGVPQK 454
QY 326 TCDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEEFGKLR 385
DB 455 ICHMLGLAELGTVCSS--SSSCSIVQDTGLPTAFTMAHELGHILNMNDDDDKCMYVTRON 513
QY 386 AN---HMSPTL--IQDRANPWSACSAIITDPLDSGHGDCILLDQSPKPSLSP---EDLP 438
DB 514 NNKVLHMSVMGIHM---HPWSNKK---TD-----KSCL--ETISVCAHIPYGTLP 559
QY 439 GASVYLSQOCELAFGVSGKPCPYMXYCTKLWC---TGKAKGMVQCOTRHPWADGTSCG- 494
DB 560 GEIYSLDAQQLSFGNDFGYPCTDECKRLWCNRTSGNSNEQ--CASSNLPWADGTPCGS 617
QY 495 EGKCLKGACVERHNLNKH-----RVDSWAKWDPYPCSCTCGGVQLARRCTNTPAN 550
DB 618 SGHWCRGKCVS---NKHGYGRQVNGMGWGPWTPFTPCSLTCGGVQESRRRECQVDPEN 673
QY 551 GGYCEGVRYKRSNLEPCPSSASGKSFREOCEAFNGYNHSTNRLTLAVAVWPYKISGV 610
DB 674 GGYCTGSRKRYKSCNTHQCPGSMDB--REQOCYAMNGRNMNIPGVNPDTKWVPKYE-- 729
QY 611 SPDRCKKLICRANGTYFYVYVLPKVVVDGTLCSPDSTSVCVQKCIKAGCDGNLGSKKRFD 670
DB 730 --KDACKLFCRDMKVITYFMLKSNVTDGTSCAVDSFDKCVNGICRPAGCDNELNSIAKLD 787
QY 671 KCGVCGDKNCKKVTG 687
DB 788 KCGVCEGRNDTCHEVTG 804

RESULT 8
Q8TE57 PRELIMINARY; PRT; 1072 AA.
AC Q8TE57
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Metalloprotease disintegrin 16 with thrombospondin type I motif.
CN ADAMTS16.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE-21856482; PubMed-11867212;
RA Cai S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RT Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
seven novel human ADAMTSs, a family of metalloproteinases with
disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
DR EMBL; AJ315734; CAC86015.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1072 AA; 119491 MW; 081EEFF78F47D061 CRC64;

Query Match 25.8%; Score 967.5; DB 4; Length 1072;
Best Local Similarity 35.1%; Pred. No. 1.1e-71;
Matches 243; Conservative 86; Mismatches 261; Indels 103; Gaps 24;

QY 51 QGLIFQIFATQEDFYHLPTDAQFLAPAFSTEHLGVPLQGLTGSS-----DLRCFY 103
DB 95 ESLHLRKGPRHDFHMDLRTSSSLVAPGFIVQTL-----GKTGTSVQTLPPEDF--CFY 147
QY 104 SGDYNAPDPSFAVSLCGGLRGAFYGAEEYISPLPNASAPAAQRNSQ-----AHLQRR 160
DB 148 QGSLSRHNSVALSTCGLSGMTIRTEADYFLRLPLSHLSWKLGRAAQSSFSHVLYKR 207
QY 161 G---VPGGPGSDPTSR-----CGVASGNPA-----ILRA 187
DB 208 STEHPAGASVLTSTWELAHQPLHSSDLRLGLPQKHFCGRKKYMPQPKEDLFL 267
QY 188 LDPYKPRNAGFESRRSRGRKRFVSIPIRYETLVVADESMVKFNG-ADLEHYLLTL 246
DB 268 PDEYK---SCLRHKSLRLSRNEEL-----NVTLVVDKMKMNQHNENITVLTIL 319
QY 247 ATAARLYRHPISILNPIVIVVKLLLRDRSGPKVTGNAALTLRNFCWQKLNKYS DK 305
DB 320 NNVSALFKDGTIGGNINAIIVGLILLEDEQPLVISHADHTLSSFCOMOSGLMGKQTR 379
QY 306 HPEYWDYTAILFTRODLG--ATTCDTLGMADVGMCDPKRSCSVIEDDGLPSAFTTAHEL 363
DB 380 H-----DHAILLTGLDICSWKNEPCDTLGFAPISGMSKYSRCTINEDTGLAFTIAHES 435
QY 364 GHVFNMPHDNV-KVCEEVFGKLRANHMSPILQIDRANPWSACSAIITDPLDSGHGDC 422
DB 436 GHNFGMIHDGEGNMCKSEG-----NIMSPTLAGNNGVFSWSPCSROYLHKFLSTAQAIC 490
QY 423 LLDQSPKI---SLPEDLPASVYTLISQOCELAFGVSGKPCPY---MOYCTKLWC--TGKA 474
DB 491 LADQP-KPVKEYKYPEKLPGLYDANTOCKWQFGEKAKLMDLDFKDKKALWCHRIGRK 549
QY 475 KGMVQCOTRHPFWADGTSCGEGKCLKGACVERHNLNKHRVDGGSWAKWDPYPCSRCTGG 534
DB 550 -----CETKFPAAEGTICGHDMWCRGGQCVKYGDGPKPTHGHWSWSSWSPCSRTCGG 604
QY 535 GVQLARRCCTNTPANGKYCEGVRYKRSNLEPCPSSASGKSFREEQ-----EAFNG 589
DB 605 GVSHRSRLCTNPKPSHGKFCGEGSTRTILCNLSQKCRDPS--VDFRAAQCAEHNSRRFRG 662
QY 590 YNHSTNRLTLAVAVWPYKISGVSPDRCKKLICRANGTYFYVYVLPKVVVDGTLCSPDSTSV 649
DB 663 RHYK-----WKP-YTQVEDQDLCKLYCIAEGDFEFLSNKVKDGTPCSEDSRNV 712
QY 650 VQGCCKIAGCDGNLGSKKRFRDKCGVCGGDKNKSC 682
DB 713 IDGICERVCDNVLGSDADEVDCVGCNGNNSAC 745

RESULT 9
Q8TE59 PRELIMINARY; PRT; 1207 AA.
AC Q8TE59
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE ADAMTS-19.

ADAMTSL9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=21856482; PubMed=11867212;
RX Cal S.; Obay A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.
RT *Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.*;
RL Gene 283.49-62(2002);
DR ENBL; AJ311904; CAC84565.1; -;
SQ SEQUENCE 1207 AA; 134061 MW; AF36F6BF5886FDE2 CRC64;

Query Match 25.1%; Score 941.5; DB 4; Length 1207;
Best Local Similarity 32.3%; Pred. No. 1 9e-69;
Matches 247; Conservative 100; Mismatches 278; Indels 141; Gaps 28;

QY 16 AGGPEPREVVVTRLDPDINGRR----- 39
|| | : | : | : | : ||
Db 76 AAGSREVRSVAPLEPVEGRSESRLLPPPSGEDEELESQELPRGSSGAALSPG 135

QY 40 --YYWRGP-----EDSGDGLIFQITAFQEDFYHLHLPDAQFLAPAFSTE 82
|| : | : | : | : | : ||
Db 136 APASWQPDPPPPQQPPPPPAQHAEPDGDG-VLLRIPAFSRDLYLRLLRDRGRFLAPFAVE 194

QY 83 HLGVPLQGLTGSSDLR-----RCFYSDVNAAEPDSFAAVSLCSGLRGAFGYRGAYEV 135
|| | : | : | : | : | : ||
Db 195 QRPNPGFGPTCAASPAPPADGCGTYTGNVLRHPCSLASFSTCCGGGLGKGFQLNEDFI 254

QY 136 -ISLPLNASPAA-----QRNSQGAHLLQRRGVPGGSDPTSRCGVASGWNPAILRA 187
|| | : | : | : | : | : ||
Db 255 FIEPLNDMTAITGHPRVRYQKRSMEEKVTEK-----SALHSHYCGIIS----- 298

QY 188 LDVPKRRRAGFGESESRSSRRGRKRF--VSIPR--YVETLVVADESVMVKFHGAD-LEHYLL 243
|| | : | : | : | : | : ||
Db 299 -DKGRPR-----SRKTAESGRKARYSKYLPQEVNIETVVVADPAMVSYHGADAARFIL 351

QY 244 TLATAARLYRHPISILNPINIVVYKVVLLLRDROSGPKVTGNAALTIRNFCAWQ-KKLNV 302
|| : | : | : | : | : | : ||
Db 352 TILMVNPLFQKSLGVOVNLRIKILLHETPELYIGHGKMLSECFCKWQHEEFKK 411

QY 303 SDKHPEY--W-----DFALLFTRODLG-GATTCDTLGMADVGTMCDDPKSCSVIED 350
|| | : | : | : | : | : ||
Db 412 NDHLEMSTNWGEDMSTSVDAAILLTRKDFCHKDEPCDTVGIALYLSGCMSEKRRKIITAE 471

QY 351 DGLPSAFTTAHELCHVENPMPHDN-VKKCEVFVKLRANHMMSPTLIQIDRAN--PWSAQS 407
|| | : | : | : | : | : ||
Db 472 NGLNLAFTIAEHGHNNGINHNDHPSCAD-----CLHIMSGEWIQKNLGDVSWSRCS 525

QY 408 AAITDPLDSHGDCDLLD---QPSKPISLPEDLPFGASYTL SQCELAFAGVSGKPCPTMYQ 464
|| : | : | : | : | : | : ||
Db 526 KEDLERPLRSKASNCLLTQNPQSVNSVWPVKLPGMTYTADQCQILFGPLASFCQEMQH 585

QY 465 --CTKLWCTCKAKQMVCQTFRHPWDGTCGGGEGKCLKKACVERHNLNKHRRVDGSWAKW 522
|| | : | : | : | : | : | : ||
Db 586 VICTGLWC---KVGEKECRTKLDPDMGTDGCDLKGWK KAGECTSRTPAPEH-LAGEWSLW 642

QY 523 DPYPGCSRTCGGVQLARRCOTNPTPANGKYCEGVRVKYSRNLNLEPCPSSASGKSFREE 582
|| | : | : | : | : | : | : ||
Db 643 ---SPCSRTCAGISSPERKC--PGLDSEARDCNGRPKNQYRICENPPCAPGLPG--FROW 695

QY 583 QCFAFNYNHSTNRLLTAVAWVPKYSGVSPDRK-CKLICRANGTGYFYVLAPKVVWGDTLC 641
|| : | : | : | : | : | : ||
Db 696 QCOAYSVRTSPKHIL-----QWQAVLDEEKPCALFCSPVKEQPIILLSKVMGDTSC 748

QY 642 SPDSTSVCVQGKCIKAGCDNGLSKRRFKCGVCGGDNKSCKKVTG 687
|| : | : | : | : | : | : ||
Db 749 GYGGLDICANGRCOKYCGDLGLSLAREDHCGYCNGCNGSKCIK 794

```
Db 540 CIWKSPQTYGQDGGSSWTKFGSCSRSCGGVRSRSCNNPSPAYGGRPCLGPMFEYQ 599
QY 564 SCNLEPCPSASCKSFREOCEAFNGY-----NHSTNRLTLAVAVWPKYSGVSRDKCK 617
Db 600 VCNSEECPGTY--EDFRAQOCCAKRNSYYVHQNAKHS-----WVP-YEPDDDAQKCE 647
QY 618 LICRANGTGYFYVLAPKVVDGTLCS-PDSTSVCVQKCIKAGCDGNLGSKKRFRDKGCVG 676
Db 648 LICQSDATGVDVFMNQVVDGTRCSYRDPYSVCARGECVPVGCDEKVGSMKADDKCGVCG 707
QY 677 GDNKSKCKVTG 687
Db 708 GDNSHCRTVKG 718

RESULT 11
Q8TE55 PRELIMINARY; PRT; 1223 AA.
ID Q8TE55;
AC Q8TE55;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Metalloprotease-disintegrin protease.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FETAL LUNG;
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:45-62(2002).
RW EMBL; AJ345098; CAC87943.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1223 AA; 133925 MW; D58B56593977ED15 CRC64;

Query Match 24.8%; Score 928; DB 4; Length 1223;
Best Local Similarity 33.1%; Pred. No. 2.5e-68;
Matches 222; Conservative 95; Mismatches 276; Indels 78; Gaps 22;

QY 45 PEDSGDGLIFQITAFQEDFYHLTPDAQFLAPAFSTEH-----LGVPLQGLTGGSSD 97
Db 98 PGRVGRHSLYFNVTGKELHLRLRPNRRLVPGSSVWEQEDFRELFRQLR-----149
QY 98 LRRCFYSGDVNABPDSFAAVSLCGGLRGAFYRGAEYVISPFPNASAPAAQRNSQG-AHL 156
Db 150 -QECVYTGVTGMPGAAGAAVSNCDGLAGLIRTDSTDFEPLERQ---QEKEASGRTHV 205
QY 157 LQRRGVP-----GGPSGDPTRSGVAGSNWPAILRALDPKPRAGFESRRSRRAKRAK 212
Db 206 VYRREAVQBEAPDPDGLHNE-AFGLGDLPLNLLGLVGQ-----LGDTERKRRHAKPGS 258
QY 213 FVSIPRVVETLVVADESMVKFHCAD-LEHYLLTLATAARLYRHPHSILNPINIVVKLL 271
Db 259 Y-----SIEVLLVDDSVVRFGKEHVQNYVLTLMNIVDEIYHDSGLGVHINIALVRLIM 313
QY 272 LRDRDSGPKV-TGNAALTILNFCAMOKLNKYSKHPYWDTAIFTTRDLCGATTCDTL 330
Db 314 VGYRQSLSLIERGNPSRSLQVCRWAHSQORQDPSHAHHVFLTRQDF-----GPS 367
QY 331 GMADVGTMCDPKRSQSVIEDDGLPSAFTTAHELGHVFNMPHD---NVKVCBEVFGKLRAN 387
Db 368 GYAPVTGMCHPLRSCALNHEDGFSSAFVTAHETGHRVILGHVGDGQGCADFTSLGS----423
QY 388 HMASPTLIQIDRANPWSACSAALITFDLSDGHGDCCLDQPSKPI-SLPEDLPASVYTLQ 446
```

```
Db 424 -VMAPLQAAAFHFRHWSRCSKLELSRYLPS--YDCLDDPDPFPAWPOPPPELPGINYSMDE 480
QY 447 QCELAFAVGSGKPC--PYMOYCTKLWCTGKRAKGOMVQOTRHFWDADTSCGEGKLCCLKGA 503
Db 481 QCRDFDSGYOTCLAFRTFEPCKQLNCS-HPDNPYFCKTKKGPPLDGTGECAPGKWCFKGH 539
QY 504 CVERHNLNKHVRDWSWAKWDPYPCSRCTGGGVLARROCTNPTPANGKYCEGVRYKYR 563
Db 540 CIWKSPQTYGQDGGSSWTKFGSCSRSCGGVRSRSCNNPSPAYGGRPCLGPMFEYQ 599
QY 564 SCNLEPCPSASCKSFREOCEAFNGY-----NHSTNRLTLAVAVWPKYSGVSRDKCK 617
Db 600 VCNSEECPGTY--EDFRAQOCCAKRNSYYVHQNAKHS-----WVP-YEPDDDAQKCE 647
QY 618 LICRANGTGYFYVLAPKVVDGTLCS-PDSTSVCVQKCIKAGCDGNLGSKKRFRDKGCVG 676
Db 648 LICQSDATGVDVFMNQVVDGTRCSYRDPYSVCARGECVPVGCDEKVGSMKADDKCGVCG 707
QY 677 GDNKSKCKVTG 687
Db 708 GDNSHCRTVKG 718

RESULT 12
Q8TEY8 PRELIMINARY; PRT; 1159 AA.
ID Q8TEY8
AC Q8TEY8;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE ADAMTS14.
GN ADAMTS14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21839041; PubMed=11741898;
RA Colige A., Vandenbergh I., Thiry M., Lambert C.A., Van Beeumen J.,
RA Li S.W., Prockop D.J., Lapiere C.M., Nussgens B.V.;
RT "Cloning and Characterization of ADAMTS-14, a Novel ADAMTS Displaying
RT High Homology with ADAMTS-2 and ADAMTS-3.";
RL J. Biol. Chem. 277:5756-5766(2002).
RW EMBL; AF366351; AAL79814.1; -.
SQ SEQUENCE 1159 AA; 127336 MW; A5B130149BF7FF34 CRC64;

Query Match 24.78%; Score 926.5; DB 4; Length 1159;
Best Local Similarity 33.2%; Pred. No. 3.1e-68;
Matches 223; Conservative 96; Mismatches 277; Indels 75; Gaps 23;

QY 45 PEDSGDGLIFQITAFQEDFYHLTPDAQFLAPAFSTEH-----LGVPLQGLTGGSSD 97
Db 31 PGRVGRHSLYFNVTGKELHLRLRPNRRLVPGSSVWEQEDFRELFRQLR-----82
QY 98 LRRCFYSGDVNABPDSFAAVSLCGGLRGAFYRGAEYVISPFPNASAPAAQRNSQG-AHL 156
Db 83 -QECVYTGVTGMPGAAGAAVSNCDGLAGLIRTDSTDFEPLERQ---QEKEASGRTHV 138
QY 157 LQRRGVP-----GGPSGDPTRSGVAGSNWPAILRALDPKPRAGFESRRSRRAKRAK 212
Db 139 VYRREAVQBEAPDPDGLHNE-AFGLGDLPLNLLGLVGQ-----LGDTERKRRHAKPGS 191
QY 213 FVSIPRVVETLVVADESMVKFHCAD-LEHYLLTLATAARLYRHPHSILNPINIVVKLL 271
Db 192 Y-----SIEVLLVDDSVVRFGKEHVQNYVLTLMNIVDEIYHDSGLGVHINIALVRLIM 246
QY 272 LRDRDSGPKV-TGNAALTILNFCAMOKLNKYSKHPYWDTAIFTTRDLCGATTCDTL 330
Db 247 VGYRQSLSLIERGNPSRSLQVCRWAHSQORQDPSHAHHVFLTRQDF-GPSGMO--303
QY 331 GMADVGTMCDPKRSQSVIEDDGLPSAFTTAHELGHVFNMPHD---NVKVCBEVFGKLRAN 387
```

```
Db 304 GYAPVTGMCHPLRSCALNHEDGFSSAFVIAHETGHVLGMEHDGQNGCADETSLGS----- 359
QY 388 HMMSPTLIQIDRANPWSACSAIIITDFLDGSHGDCLLDOPSKPI-SLPEDLPASVYTLQ 446
Db 360 -VNAPLVQAFAHFRHSRCSKLSYSLPS--YDCLLDPPFPANPQPELPFGINYSWDE 416
QY 447 QCELAFGVSGKPG---PYMOYCTKLWCTGKAKQMVQOTRHPFPWADGTCGSGKLCCLKGA 503
Db 417 QCRFDGSGYQTCLAFRTFEPCKQLWCS-HPDNPYFCKTKGPPDLDTBCAPGKWCXKG 475
QY 504 CVERHNLNKHVRDGSNAKWDYPGCSRCTCGGGVQLARRQCTNPTPANGKCYCEGVRYK 563
Db 476 CIWKSPEQYTGQDGWSSWTKFGSCSRGCVRSRSCNPNPSPAYGGRPCLGPMFEYQ 535
QY 564 SCNLEPCSSASGKSFREEQCAFNGY-----NHSTNRLTLAVAWVPKYSGVSPDKCK 617
Db 536 VCNSEECPGTY--EDFRAQCAKRNYSYVHQNAKHS-----WVP-YEPDDDAQKCE 583
QY 618 LICRANGTGYFYVLAIPKVVDTGLCS-PDSTSVCVQKCIKAGCDGNLGSKKRFDKCGVG 676
Db 584 LICQSDTGDVFMNVOVHDGTRCSYRDPYSVCARGCEVPVGVGDKVEGSMKADDRKCGVG 643
QY 677 GDNKSCCKVTG 687
Db 644 GDNSHCRTVKG 654

RESULT 13
Q9GL54 PRELIMINARY; PRT; 269 AA.
ID Q9GL54
AC Q9GL54;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Aggrecanase-2 (Fragment).
GN ADAMTS-11.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP Goad D.L., Goad M.E.;
RT "Molecular cloning of lapine aggrecanase-2 (ADAMTS-11) from articular
RL Submitted (OCN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF317415; AAC33062.1; -.
DR HSSP; Q9PW35; IBDU.
DR MEROPS; M12.225; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp-1; 1.
DR SMART; SM00209; TSPl; 1.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSPl; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1
FT SIGNAL 1
FT CHAIN 269
SQ SEQUENCE 269 AA; 29193 MW; 97A1CA80B33452FA CRC64;
```

```
Query Match 22.2%; Score 830; DB 6; Length 269;
Best Local Similarity 54.7%; Pred. No. 4.6e-61;
Matches 146; Conservative 37; Mismatches 7; Indels 10; Gaps 3;

QY 329 TLGMADVTGTCMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVEFGKLRANH 388
Db 1 TLGMADVTGTCSPERSCAVIEDDGLHAFTVAHEIGHLLGLSHDSDSKFCEENFGSTEDKR 60

QY 389 MMSPTLIQIDRANPWSACSAIIITDFLDGSHGDCLLDQPSKPISLPDLPGASYTLQ 448
Db 61 LMSSILTSIDASKPWSKCTSATITEFLDDHGNCLLDVPKQILGPEELPGQTYDATQC 120
```

```
QY 449 ELAFGVGSKPCPYMOYCTKLWCTGKAKQMVQOTRHPFPWADGTCGSGKLCCLKGACVER- 507
Db 121 NLTFGEPTVCPGMDVCARLWCAVVRQGMVCLTKKLPAVEGTGCGKRICLOGKCVDKT 180
QY 508 ----HNLNKHVRDGSNAKWDYPGCSRCTCGGGVQLARRQCTNPTPANGKCYCEGVRYK 563
Db 181 KKTYSTSSH---GNWGSWGPWGQCSRSCGGVQFAYRHCNPNAPRNSGRYCTGKRAIYR. 237
QY 564 SCNLEPCSSASGKSFREEQCAFNGY 590
Db 238 SCSVTPCP--ANGKSFREHQCEARNGY 262

RESULT 14
Q96L37 PRELIMINARY; PRT; 1427 AA.
ID Q96L37
AC Q96L37;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE von Willebrand factor-cleaving protease precursor.
GN ADAMTS13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX PubMed-11557746;
RA Zheng X., Chung D., Takayama T.K., Majerus E.M., Sadler J.E.,
RA Fujikawa K.;
RT "Structure of von Willebrand Factor-cleaving Protease (ADAMTS13), a
RL Metalloprotease Involved in thrombotic Thrombocytopenic Purpura.";
RL J. Biol. Chem. 276:41059-41063(2001).
DR EMBL; AY055376; AAL17652.1; -.
DR MEROPS; M12.241; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPl.
DR InterPro; IPR000130; Zn_Mtpeptdse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp-1; 4.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50092; TSPl; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Protease; Signal.
FT SIGNAL 1
FT CHAIN 75
FT SIGNAL 33
FT CHAIN 1427
SQ SEQUENCE 1427 AA; 153632 MW; EB1BC3AABC1A4442 CRC64;
```

```
Query Match 19.2%; Score 720; DB 4; Length 1427;
Best Local Similarity 31.5%; Pred. No. 6.8e-51;
Matches 179; Conservative 66; Mismatches 240; Indels 84; Gaps 18;
```

```
QY 174 CVASG-WNP-----AILRALDPY-----KPRRAGFESRRSRSSRAKRF 213
Db 21 CGFLLCWGPESHFQSCLOALEPQAVSYLSFGAPLKGPPSPGFQQRQQR-----RA 75
QY 214 VSIPIRYETLVVADESVMYKFGADLEHYLLTLATAARLYRHPHSILPINIVVVKVLLLR 273
Db 76 AGGILHELLEVAVGPDVFOAHOEDTERVLTNLNIGAELLRDPISLGAFRVHLVKMVLIT 135
QY 274 DRDGPVKVTGNAALTLRNFCAWOKKLNKVSDEKHPYWDTAILFTRQDL-CGATTCDFLGM 332
Db 136 EPEGAPNTANLTSSLISVCWGSQTINPEDDTDFGHADLVLYITRFDLPDGNRQVRG 195
QY 333 ADVGTMCDPKRSQSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVEFGKLRANHMSP 392
Db 196 TQLGACGSPVTSCLITETGFDLGVTAHEIGHSGFLEHGDGAPGS----CGGPSCHVNAS 251
QY 393 TLIQIDRANP-----WSACSAIIITDFLDGSHGDCLLD-----QPSKPISLPDLPGASYT 443
```

```
Db 252 -----DGAAPRAGLANSPCSRROLLSLLSAGRARCVMDPPRPQPGSAGHPPDAQPGLIYS 306
QY 444 LSQOCELAFGVGSKPCPY-----MOYTKLWCTGKAKGMVQCOTRHPFPWADGTSCGEGKLC 499
Db 307 ANEQCRVAFGPKAVACTFAREHLDMQALSCHTDPLDQSSCSRLLVPLLDGTGCGVEKWC 366
QY 500 LKAGACVERHNLNK-HRVDSWAKWDYGPCSRCTCGGQVLARRQCTNPTPANGKYCEGV 558
Db 367 SKGRCSLVELTPIAAVHGRWSSWGSPRSCSRSCGGVTVRRQCNPRPFAFGACVGA 426
QY 559 RVKYRSCNLEPCPSSASGSKFRECEAFNG-----YNHSTNRLTLAVAW---V 604
Db 427 DLAEMCNTQACEKTO--LEFMSQQCARDTGOPLRSSPGGASFYH-----WGAAY 474
QY 605 PKYSGVSPDKCKLICRANGTYFYVYLPKVVDTGLCSP-----DST-SVCVQGGKIKAG 658
Db 475 PHSQGDAAA--LCRHMCRAIGESIMKRGDSFLDTRCMPSPGPDGDTLSLCVSGSCRTFG 531
QY 659 CDGNLGSKKRFDCGVCYGGDNKCKKVTG 687
Db 532 CDGRMDSQQVWDRQCQVCGGDNSTCSPRK 560

RESULT 15
Q95N24 PRELIMINARY; PRT; 192 AA.
AC Q95N24;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Aggrecaanase-1 (Fragment).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RA Flannery C.R., Little C.B.;
RT "Expression and activity of equine aggrecaanases.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AF368321; AAK53425.1; -.
DR MEROPS; M12.221; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000130; Zn_Mtpeptase.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 192
SQ SEQUENCE 192 AA; 20670 MW; 9013B0E19FCE8C56 CRC64;

Query Match 14.9%; Score 558; DB 6; Length 192;
Best Local Similarity 51.0%; Pred. No. 1.4e-38;
Matches 98; Conservative 26; Mismatches 64; Indels 4; Gaps 3;

QY 332 MADVGTMDCKPSCSVTEDDGLPSAFTTAHELGHVFNMPHDNVKCEVEFGK-LRANHMM 390
Db 1 MADVGTVDPCARCAIVEDDGLQSATAHELGHVFNMLHDNSKPCVGLNGPGSGSTRHYM 60
QY 391 SPTLIQIDRANPWSACSAALITDLSGHGDCILLDFQSKPISLPDLPGASYTLSSQOCEL 450
Db 61 APVMAHVDPEEPWSPCSARFITDLDNGYGHCLLDKPEAPLHLPVTFPGKYDADRCQCL 120
QY 451 AFGVGSKPCPYM-QYCTKLWCTGKAKGMVQCOTRHPFPWADGTSCGEGKLC LKAGACVERHN 509
Db 121 TFGPDSRHCPQLPPPCAAALWCSGLHNGHAMCOTKHSFPWADGTPCGPAQACMGRCCLHMDQ 180
QY 510 LNKHRVD--GSW 519
Db 181 LQEFNIPQAGW 192
```

Search completed: April 29, 2003, 17:17:53
Job time : 45.8928 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:12:21 ; Search time 12.9594 Seconds
(without alignments)
3604.758 Million cell updates/sec

Title: US-10-009-332-1-copy_1_583

Perfect score: 3163

Sequence: 1 MLLGLITLAFAGRTAGGFE.....SCNLEPCSSAGSKSFREQ 583

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- PublishedApplications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3155	99.7	950	10 US-09-965-631-4	Sequence 4, Appli
2	3126.5	98.8	823	9 US-10-163-316-2	Sequence 2, Appli
3	1913	60.5	367	10 US-09-965-631-6	Sequence 6, Appli
4	1669	52.8	321	10 US-09-965-631-2	Sequence 2, Appli
5	1562.5	49.4	950	10 US-09-321-987B-4	Sequence 4, Appli
6	1562.5	49.4	968	9 US-10-163-316-7	Sequence 7, Appli
7	1560.5	49.3	967	12 US-10-105-929-2	Sequence 2, Appli
8	1353	42.8	727	9 US-10-097-597-12	Sequence 12, Appli
9	1353	42.8	727	9 US-10-097-580-12	Sequence 12, Appli
10	1353	42.8	727	10 US-09-445-023A-12	Sequence 12, Appli
11	1351	42.7	727	9 US-10-097-597-1	Sequence 1, Appli
12	1351	42.7	727	9 US-10-097-580-1	Sequence 1, Appli
13	1351	42.7	727	10 US-09-445-023A-1	Sequence 1, Appli
14	1337.5	42.3	837	9 US-10-174-590-352	Sequence 352, App
15	1337.5	42.3	837	9 US-10-176-758-352	Sequence 352, App
16	1337.5	42.3	837	9 US-10-175-737-352	Sequence 352, App
17	1337.5	42.3	837	9 US-10-173-706-352	Sequence 352, App
18	1337.5	42.3	837	9 US-10-175-738-352	Sequence 352, App
19	1337.5	42.3	837	9 US-10-175-752-352	Sequence 352, App

20	1337.5	42.3	837	9 US-10-176-482-352	Sequence 352, App
21	1337.5	42.3	837	9 US-10-176-757-352	Sequence 352, App
22	1337.5	42.3	837	9 US-10-176-913-352	Sequence 352, App
23	1337.5	42.3	837	9 US-10-180-552-352	Sequence 352, App
24	1337.5	42.3	837	9 US-10-180-557-352	Sequence 352, App
25	1337.5	42.3	837	9 US-10-173-700-352	Sequence 352, App
26	1337.5	42.3	837	9 US-10-174-572-352	Sequence 352, App
27	1337.5	42.3	837	9 US-10-174-579-352	Sequence 352, App
28	1337.5	42.3	837	9 US-10-174-582-352	Sequence 352, App
29	1337.5	42.3	837	9 US-10-174-588-352	Sequence 352, App
30	1337.5	42.3	837	9 US-10-175-739-352	Sequence 352, App
31	1337.5	42.3	837	9 US-10-175-740-352	Sequence 352, App
32	1337.5	42.3	837	9 US-10-175-743-352	Sequence 352, App
33	1337.5	42.3	837	9 US-10-176-488-352	Sequence 352, App
34	1337.5	42.3	837	9 US-10-176-492-352	Sequence 352, App
35	1337.5	42.3	837	9 US-10-176-747-352	Sequence 352, App
36	1337.5	42.3	837	9 US-10-176-750-352	Sequence 352, App
37	1337.5	42.3	837	9 US-10-176-985-352	Sequence 352, App
38	1337.5	42.3	837	9 US-10-176-987-352	Sequence 352, App
39	1337.5	42.3	837	9 US-10-176-991-352	Sequence 352, App
40	1337.5	42.3	837	9 US-10-176-992-352	Sequence 352, App
41	1337.5	42.3	837	9 US-10-176-993-352	Sequence 352, App
42	1337.5	42.3	837	9 US-10-184-658-352	Sequence 352, App
43	1337.5	42.3	837	9 US-10-173-695-352	Sequence 352, App
44	1337.5	42.3	837	9 US-10-173-697-352	Sequence 352, App
45	1337.5	42.3	837	9 US-10-173-705-352	Sequence 352, App

ALIGNMENTS

RESULT 1
US-09-965-631-4
; Sequence 4, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Enco
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 950
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-965-631-4

Query Match 99.7%; Score 3155; DB 10; Length 950;
Best Local Similarity 99.8%; Pred. No. 9.9e-251;
Matches 582; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MLLLGLITLAFAGRTAGGPEPEREVVPLRLDPDNLNGRRYYWRGPDSDGQGLIFQITAF	60
Db	1	MLLLGLITLAFAGRTAGGPEPEREVVPLRLDPDNLNGRRYYWRGPDSDGQGLIFQITAF	60
QY	61	QEDFYHLTPDQAQFLAPAFSTHGLGVPLQGLTGGSSDLRRRCFYSGDVNAEPDSFAVSILC	120
Db	61	QEDFYHLTPDQAQFLAPAFSTHGLGVPLQGLTGGSSDLRRRCFYSGDVNAEPDSFAVSILC	120
QY	121	GGLRGAFYRGAEYVISPLPNASAPAAQNSOGAHLQRRGVPGPGSDPTSRCSVASGW	180
Db	121	GGLRGAFYRGAEYVISPLPNASAPAAQNSOGAHLQRRGVPGPGSGDPTSRCSVASGW	180
QY	181	NPAILLALDPYKPRAGFESRRSRGRKRFVSIPIRYVETLVVADFSWKFGHADLH	240
Db	181	NPAILLALDPYKPRAGFESRRSRGRKRFVSIPIRYVETLVVADFSWKFGHADLH	240
QY	241	YLLTLTAAARLYRHPSILNPINIVVVKVLLLRDRDSGPKVTGNAALTNRNCAWQKLN	300

|||||
Db 241 YLLTLATAARLYRHPSTILNINIVVVKLLLRDRSDGPKVTGNAALTLRNFCWQKKLN 300
QY 301 KVS DKHPEYWDTAILFTTRQDLCGATTCDTLGMADVGTMCDPKRS CVIEDDGLPSAFTTA 360
Db 301 KVS DKHPEYWDTAILFTTRQDLCGATTCDTLGMADVGTMCDPKRS CVIEDDGLPSAFTTA 360
QY 361 HELGHVFNMPHDNKKVCEEVFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSHG 420
Db 361 HELGHVFNMPHDNKKVCEEVFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSHG 420
QY 421 DCLLDQPSKPISLPEDLPAGASYTLSQOCELAFGVGSKPCPYMQYCTKLMCTGKAKGMVC 480
Db 421 DCLLDQPSKPISLPEDLPAGASYTLSQOCELAFGVGSKPCPYMQYCTKLMCTGKAKGMVC 480
QY 481 QTRHFPWADGTSCEGKLCILKGACVERHNLNKRVDSWAKWDPYGPCSRTCGGGVQLAR 540
Db 481 QTRHFPWADGTSCEGKLCILKGACVERHNLNKRVDSWAKWDPYGPCSRTCGGGVQLAR 540
QY 541 RQCTNPTPANGKYCEGVYKRSCLNLEPCPSSASGKSFREQ 583
Db 541 RQCTNPTPANGKYCEGVYKRSCLNLEPCPSSASGKSFREQ 583

RESULT 2
US-10-163-316-2
; Sequence 2, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MPI01-025PIRNM
; CURRENT APPLICATION NUMBER: US/10/163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-316-2

Query Match 98.8%; Score 3126.5; DB 9; Length 823;
Best Local Similarity 95.9%; Pred. No. 1.8e-248;
Matches 581; Conservative 0; Mismatches 2; Indels 23; Gaps 1;
QY 1 MLLLGILTLAPAGTAGGFEPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
Db 1 MLLLGILTLAPAGTAGGFEPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFCYSGDVNAEPDSPAASVLC 120
Db 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFCYSGDVNAEPDSPAASVLC 120
QY 121 GGLRGAFGYRGAEEYVISPPLNAPASAPAAQNSQGAHLQRRGVPGSGDPTSRCGVASGW 180
Db 121 GGLRGAFGYRGAEEYVISPPLNAPASAPAAQNSQGAHLQRRGVPGSGDPTSRCGVASGW 180
QY 181 NPAILRALDPYKPRRAGFGESESRSSRRGRKRFVSIPIRYVETLVVADSVKFKHGADLEH 240
Db 181 NPAILRALDPYKPRRAGFGESESRSSRRGRKRFVSIPIRYVETLVVADSVKFKHGADLEH 240
QY 241 YLLTLATAARLYRHPSTILNINIVVVKLLLRDRSDGPKVTGNAALTLRNFCWQKKLN 300
Db 241 YLLTLATAARLYRHPSTILNINIVVVKLLLRDRSDGPKVTGNAALTLRNFCWQKKLN 300
QY 301 KVS DKHPEYWDTAILFTTRQDLCGATTCDTLGMADVGTMCDPKRS CVIEDDGLPSAFTTA 360
Db 301 KVS DKHPEYWDTAILFTTRQDLCGATTCDTLGMADVGTMCDPKRS CVIEDDGLPSAFTTA 360

QY 361 HELGHVFNMPHDNKKVCEEVFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSHG 420
Db 361 HELGHVFNMPHDNKKVCEEVFGKLRANHMSPPTLIQIDRANPWSACSAAIITDFLDSHG 420
QY 421 DCLLDQPSKPISLPEDLPAGASYTLSQOCELAFGVGSKPCPYMQYCTKLMCTGKAKGMVC 480
Db 421 DCLLDQPSKPISLPEDLPAGASYTLSQOCELAFGVGSKPCPYMQYCTKLMCTGKAKGMVC 480
QY 481 QTRHFPWADGTSCEGKLCILKGACVERHNLNKRVDSWAKWDPYGPCSRTCGGGVQLAR 540
Db 481 QTRHFPWADGTSCEGKLCILKGACVERHNLNKRVDSWAKWDPYGPCSRTCGGGVQLAR 540
QY 518 SWAKWDPYGPCSRTCGGGVQLARRQCTNPTPANGKYCEGVYKRSCLNLEPCPSSASGK 577
Db 541 SWAKWDPYGPCSRTCGGGVQLARRQCTNPTPANGKYCEGVYKRSCLNLEPCPSSASGK 600
QY 578 SFREQ 583
Db 601 SFREQ 606
RESULT 3
US-09-965-631-6
; Sequence 6, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Fridgele, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. US20020115842A1el Human Proteases and Polynucleotides Enco
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 367
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-965-631-6

Query Match 60.5%; Score 1913; DB 10; Length 367;
Best Local Similarity 99.7%; Pred. No. 3e-149;
Matches 363; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MLLLGILTLAPAGTAGGFEPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
Db 1 MLLLGILTLAPAGTAGGFEPEPEVVPVIRLDPDINGRRYYWRGPDSDGQGLIFQITAF 60
QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFCYSGDVNAEPDSPAASVLC 120
Db 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRRCFCYSGDVNAEPDSPAASVLC 120
QY 121 GGLRGAFGYRGAEEYVISPPLNAPASAPAAQNSQGAHLQRRGVPGSGDPTSRCGVASGW 180
Db 121 GGLRGAFGYRGAEEYVISPPLNAPASAPAAQNSQGAHLQRRGVPGSGDPTSRCGVASGW 180
QY 181 NPAILRALDPYKPRRAGFGESESRSSRRGRKRFVSIPIRYVETLVVADSVKFKHGADLEH 240
Db 181 NPAILRALDPYKPRRAGFGESESRSSRRGRKRFVSIPIRYVETLVVADSVKFKHGADLEH 240
QY 241 YLLTLATAARLYRHPSTILNINIVVVKLLLRDRSDGPKVTGNAALTLRNFCWQKKLN 300
Db 241 YLLTLATAARLYRHPSTILNINIVVVKLLLRDRSDGPKVTGNAALTLRNFCWQKKLN 300
QY 301 KVS DKHPEYWDTAILFTTRQDLCGATTCDTLGMADVGTMCDPKRS CVIEDDGLPSAFTTA 360
Db 301 KVS DKHPEYWDTAILFTTRQDLCGATTCDTLGMADVGTMCDPKRS CVIEDDGLPSAFTTA 360
QY 361 HELG 364
Db 361 HELG 364

RESULT 4

US-09-965-631-2
; Sequence 2, Application US/09965631
; Patent No. US20020115842A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; TITLE OF INVENTION: Hilbun, Erin
; FILE REFERENCE: LEX-0241-USA
; CURRENT APPLICATION NUMBER: US/09/965,631
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/236,689
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 321
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-965-631-2

Query Match 52.8%; Score 1669; DB 10; Length 321;
Best Local Similarity 99.4%; Pred. No. 2.8e-129;
Matches 318; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MLLGILTLAFAGTAGGFEPEREVVPRLDPDINGRRYYWRGPGSDGGLIFQITAF 60
DB 1 MLLGILTLAFAGTAGGSEFEREVVPRLDPDINGRRYYWRGPGSDGGLIFQITAF 60
QY 61 QEDFYLHLTPAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFYSGDVNAEPDSPAAYSLC 120
DB 61 QEDFYLHLTPAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFYSGDVNAEPDSPAAYSLC 120
QY 121 GGLGAFYRGAEEVISPLPNASAPAAQNSQGAHLRLQRRGVPGGSDPTSRGCVASGW 180
DB 121 GGLGAFYRGAEEVISPLPNASAPAAQNSQGAHLRLQRRGVPGGSDPTSRGCVASGW 180
QY 181 NPAILRALDPKPRRAGGESRRSRGKRRFVSIPIRYVETLVVADESVMVKFHGADLEH 240
DB 181 NPAILRALDPKPRRAGGESRRSRGKRRFVSIPIRYVETLVVADESVMVKFHGADLEH 240
QY 241 YLLTLLATAARLYRHPSTLNPINIVVKKVLLLRDRDGGPKVTGNAALTNRNFCAMQKKLN 300
DB 241 YLLTLLATAARLYRHPSTLNPINIVVKKVLLLRDRDGGPKVTGNAALTNRNFCAMQKKLN 300
QY 301 KVS DKHPEYWDTAILFTROD 320
DB 301 KVS DKHPEYWDTAILFTROE 320

RESULT 5

US-09-321-987B-4
; Sequence 4, Application US/09321987B
; Patent No. US20020102210A1
; GENERAL INFORMATION:
; APPLICANT: Kimble, Judith E
; APPLICANT: Bielloch, Robert H
; TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
; FILE REFERENCE: 960296.95386
; CURRENT APPLICATION NUMBER: US/09/321,987B
; CURRENT FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: 60/087,170
; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 60/129,023
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Murine

US-09-321-987B-4

Query Match 49.4%; Score 1562.5; DB 10; Length 950;
Best Local Similarity 50.4%; Pred. No. 6.3e-120;
Matches 315; Conservative 79; Mismatches 154; Indels 77; Gaps 16;

QY 1 MLLGILTLAFAGTAGG--FEPERVVVPRLDPDINGRRYYWRGPE-EDSGGQGLIFQI 57
DB 20 LLLASTITMLLCARGAHGRTEDEELVLP-SLE-----RAPGHDSTTTRL--RL 66
QY 58 TAFQEDFYLHLTPDAQFLAPAFSTEHLGVPLQGLTGGS-----SDLRRCFYSGDVNA 109
DB 67 DAFGQQLHLKLPDPSGFLAPGFTLQIV-----CRSGESEAQHLDPTGDLAHCEYSTVNG 121
QY 110 EPDSFAAVSLCGGLRGAFGYRGAEEVISPLPNAS-----APAAQNSQGA-----HLQRRG 161
DB 122 DPGSAAALSCEGVGAFYLGQEEFTIQAPAGVATERLAPAVPEEBSARPQPHILRRR- 180
QY 162 VPGGPGSDPTSRGCVASGWNPAILRALDPYKPRRAGFGESRRR----- 206
DB 181 -----RRSGGAKCGVMD-----DETLPYSDSRPESQNTNRNOMPVRDPTPDAGKP 226
QY 207 -----SGRAKRFVSIPIRYVETLVVADESVMVKFHGADLEHYLLTLLATAARLYRHPSTLNP 262
DB 227 SGPGRIRKKRFVSSPRYVETMLVADQSMADFGSGGLKHVLLTLFSVAARFYKHPSTIRSI 286
QY 263 NIIVVKKVLLLRDRDGGPKVTGNAALTNRNFCAMQKKLNKVS DKHPYWDTAILFTRODLC 322
DB 287 SLVVVKILVIYERQKGEPTSNAAALTNRNFCAMQKOHNSPDRDPHEYDTAILFTRODLC 346
QY 323 GATTCOTLGMADVGTWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNMPHNDVNVKCEVEVG 382
DB 347 GSHTCOTLGMADVGTWCDPSRCSVIEDDGLQAAFTTAHELGHVFNMPHNDVNVKCEVEVG 406
QY 383 KLRANHMSPTLIQIDRANPWSACSAIITDFLDSHGDCLLDQPSKPSLSPEDLPASV 442
DB 407 VTGDSHLMASLSSLDHSPQSPCSAYMTSFLDNGHGECLMDKPNQIKLPSDLPGTLY 466
QY 443 TISOQCELAFGVSKPCP-YMQYCTKLWCTGKAKGOMVCOTRHFHWADGTSCEGSKLCLK 501
DB 467 DANRQCOFTFGEESKHCDDAASCTTTLWCTGTSGGLLVCOQTKHFPWADGTSCEGSKWCVS 526
QY 502 GACVERHNLNKH---RVDGSAKWDPYGPCSRTCGGCVQLARRQCTNPTNPANGGYCEGV 558
DB 527 GKCVNKTDM-KHFATPVHSGWGPWGPWDCSRTCCGGGVGYTHRECDNPVFKNGGKCEGK 585
QY 559 RVKYSRNCLEPCSSASGSKSFREQ 583
DB 586 RVKYSRNCIEDCPDN-NGKTFREQ 609

RESULT 6

US-10-163-316-7
; Sequence 7, Application US/10163316
; Publication No. US20020197703A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 65552, A Human Matrix Metalloproteinase and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: MPI01-025PIRNM
; CURRENT APPLICATION NUMBER: US/10/163,316
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 60/297,863
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 968
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-163-316-7

Query Match 49.4%; Score 1562.5; DB 9; Length 968;

Query Match 49.3%; Score 1560.5; DB 12; Length 967;
Best Local Similarity 51.6%; Pred. No. 9.4e-120;
Matches 320; Conservative 73; Mismatches 160; Indels 67; Gaps 17;

Qy 1 MLLGILTLAFAGTAGGFEPEERVVVPIRLDPDINGRRYWRGPDGQGLIFQITAF 60
Db 36 LLLAALAVSDALGRPSEDEELVVP-ELE-----RAP---GHGTRLRHAF 81
Qy 61 QEDFVHLHTLPAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRRCYSGDVNAEPDSF 114
Db 82 DOQLDLRLPSSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCFSYCTVNGDPSSA 136
Qy 115 AAVSLCGGLRGAFYRGAEYVISLPNAS-----APAAQRNSQGA-----HLLQ--RRGYPGG 165
Db 137 AALSCEGVRGAFYILGEAYFIQPLPAASERLATAAPEKPPAPLQPHLLRRNQGVGG 196
Qy 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPYKPRRAGFSGSRRRS 207
Db 197 TCGVVDDPRPTGKAETEDGEGEPQNS-----PODPALQGVGP-TGTGS 247
Qy 208 GRAKRFVSIPIRYVETLVVVADESVMKFGADLEHYLLTLATAARLYRHPISILNPI 267
Db 244 SFGSIRKKRFVSPRYVETMLVADQSMADFHGSLKHYLLTLFSAARFYKHPISIRSI 303
Qy 263 NIVVVKLLLRDRSGPKVTGNAALTLRNFCAWQKLNKVSCKHPEYWDTAILFTRODLG 322
Db 304 SLVVVKILVIYEQKGPEVTSNAALTLRNFCAWQKLNKVSCKHPEYWDTAILFTRODLG 363
Qy 323 GATTCDFLGADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHNDNVKVCVEVFG 382
Db 364 GSHTCDFLGADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHNDNVKVCVEVFG 423
Qy 383 KLRANHMSPTLIQIDRANPWSACSAIITDFLDSGHGCLLDQPSKPSLSPEDLPASVTSQ 447
Db 424 VTGDSHLMASLSDHSQSPWPCSAIYMTSFLDNGHGECLMDKPNQIKLPDLPTLY 483
Qy 443 TLSQCELAFTGVGSKPCP-YMQYCTKLWCTGKAGQMVCTRHFPWADGTCGEGKLC 501
Db 484 DANRQCFTTEGEEKHCPDAASTCTSLWCTGSGVLVCTKHFPWADGTCGEGKLC 543
Qy 502 GACVYRNLNKH---RVDSGWAKWDPYGPCSRCTCGGVQVLARRQCTNPTPANGKYCEG 558
Db 544 GKCVNKTDM-KHFPATPVHSGWPGWPGWDCSRTCGGVQVYTMRECDNPVKNKGKCEG 602
Qy 559 RVKYRNLNKH---RVDSGWAKWDPYGPCSRCTCGGVQVLARRQCTNPTPANGKYCEG 583
Db 603 RVYRNLNKH---RVDSGWAKWDPYGPCSRCTCGGVQVYTMRECDNPVKNKGKCEG 626

RESULT 7
US-10-105-929-2
; Sequence 2, Application US/10105929
; Patent No. US20020137142A1
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Goodearl, Andrew D.J.
; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
; FILE REFERENCE: 09404/041001
; CURRENT APPLICATION NUMBER: US/10/105,929
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/130,491
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-07
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,108
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/054,961
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-105-929-2

Query Match 49.3%; Score 1560.5; DB 12; Length 967;
Best Local Similarity 51.6%; Pred. No. 9.4e-120;
Matches 320; Conservative 73; Mismatches 160; Indels 67; Gaps 17;

Qy 1 MLLGILTLAFAGTAGGFEPEERVVVPIRLDPDINGRRYWRGPDGQGLIFQITAF 60
Db 36 LLLAALAVSDALGRPSEDEELVVP-ELE-----RAP---GHGTRLRHAF 81
Qy 61 QEDFVHLHTLPAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRRCYSGDVNAEPDSF 114
Db 82 DOQLDLRLPSSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCFSYCTVNGDPSSA 136
Qy 115 AAVSLCGGLRGAFYRGAEYVISLPNAS-----APAAQRNSQGA-----HLLQ--RRGYPGG 165
Db 137 AALSCEGVRGAFYILGEAYFIQPLPAASERLATAAPEKPPAPLQPHLLRRNQGVGG 196
Qy 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPYKPRRAGFSGSRRRS 207
Db 197 TCGVVDDPRPTGKAETEDGEGEPQNS-----PODPALQGVGP-TGTGS 247
Qy 208 GRAKRFVSIPIRYVETLVVVADESVMKFGADLEHYLLTLATAARLYRHPISILNPI 267
Db 244 SFGSIRKKRFVSPRYVETMLVADQSMADFHGSLKHYLLTLFSAARFYKHPISIRSI 303
Qy 263 NIVVVKLLLRDRSGPKVTGNAALTLRNFCAWQKLNKVSCKHPEYWDTAILFTRODLG 322
Db 304 SLVVVKILVIYEQKGPEVTSNAALTLRNFCAWQKLNKVSCKHPEYWDTAILFTRODLG 363
Qy 323 GATTCDFLGADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHNDNVKVCVEVFG 382
Db 364 GSHTCDFLGADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHNDNVKVCVEVFG 423
Qy 383 KLRANHMSPTLIQIDRANPWSACSAIITDFLDSGHGCLLDQPSKPSLSPEDLPASVTSQ 447
Db 424 VTGDSHLMASLSDHSQSPWPCSAIYMTSFLDNGHGECLMDKPNQIKLPDLPTLY 483
Qy 443 TLSQCELAFTGVGSKPCP-YMQYCTKLWCTGKAGQMVCTRHFPWADGTCGEGKLC 501
Db 484 DANRQCFTTEGEEKHCPDAASTCTSLWCTGSGVLVCTKHFPWADGTCGEGKLC 543
Qy 502 GACVYRNLNKH---RVDSGWAKWDPYGPCSRCTCGGVQVLARRQCTNPTPANGKYCEG 558
Db 544 GKCVNKTDM-KHFPATPVHSGWPGWPGWDCSRTCGGVQVYTMRECDNPVKNKGKCEG 602
Qy 559 RVKYRNLNKH---RVDSGWAKWDPYGPCSRCTCGGVQVLARRQCTNPTPANGKYCEG 583
Db 603 RVYRNLNKH---RVDSGWAKWDPYGPCSRCTCGGVQVYTMRECDNPVKNKGKCEG 626

RESULT 8
US-10-097-597-12
; Sequence 12, Application US/10097597
; Publication No. US20030022352A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiji
; APPLICANT: Hakezaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: pharmaceutical composition and method of immunologically analyzing human ADA
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.0

Db 189 LMASMLSSLDHSPWSPCSAYWVTFSLDNGHGECLMDKPNPIKLPDLPGLTYDANRQC 248
QY 449 ELAFVGVSKPCP-YMOYCTKLMCTKAKQOMVQTRHFPWADGTCGEGKLCCLKGACVER 507
Db 249 QFTFGESKPCPDAATCTTLCCTGTSGLLVLCQTKHFPWADGTCGEGKWCVSGKCVNK 308
QY 508 HNLNKH---RVDSWAKWDPYPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVVRKYRS 564
Db 309 TDM-KHFATPVHSGWGPWPGWDCSRTCGGGVQVYTMRECDNPVKNKGKCEGKRVYRS 367
QY 565 CNLEPCPSSASGKSFRREQ 583
Db 368 CNLEDCPDN-NGKTFREEQ 385
RESULT 11
US-10-097-597-1
; Sequence 1, Application US/10097597
; Publication No. US20030022352A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakezaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same,
; TITLE OF INVENTION: pharmaceutical
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,597
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-597-1
Query Match 42.7%; Score 1351; DB 9; Length 727;
Best Local Similarity 63.6%; Pred. No. 1e-102;
Matches 241; Conservative 50; Mismatches 82; Indels 6; Gaps 4;
QY 209 RAKRFVSIPRYVETLVVADESVMKFGADLEHYLLTLLATAARLYRHPSTILNINIVVK 268
Db 9 RKKRFVSSPRVYETMLVADQSMAEFHGSLKHYLLTLFSAARLYKHPSTIRNSVSLVVK 68
QY 269 VLLLRDRDSGPKVTGNAALTLRNFCANQKLNKVSDDKHPEYWDTAIFTRODLGATTC 328
Db 69 ILVHDEQKGPVTSNAALTLRNFCNQKQHNPPSDRDAEHYDTAIFTRODLGSGTCD 128
QY 329 TLGMADVGTMDKPRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRANH 388
Db 129 TLGMADVGTVCPSRSCSVIEDDGLQAFTTAHELGHVFNPHDDAKQKASLNGVND 188
QY 389 MMSPTLIQIDRANPWSACSAAITDLDGSHGDCLLDQSPKPSLPEDLPGASVTL 448
Db 189 MMASMLSNLDHSPWSPCSAYWVTFSLDNGHGECLMDKPNPIKLPDLPGLTYDANRQC 248
QY 449 ELAFVGVSKPCP-YMOYCTKLMCTKAKQOMVQTRHFPWADGTCGEGKLCCLKGACVER 507
Db 249 QFTFGESKPCPDAATCTTLCCTGTSGLLVLCQTKHFPWADGTCGEGKWCVSGKCVNK 308
QY 508 HNLNKH---RVDSWAKWDPYPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVVRKYRS 564
Db 309 TDM-KHFATPVHSGWGPWPGWDCSRTCGGGVQVYTMRECDNPVKNKGKCEGKRVYRS 367
QY 565 CNLEPCPSSASGKSFRREQ 583
Db 368 CNLEDCPDN-NGKTFREEQ 385
RESULT 13
US-09-445-023A-1
; Sequence 1, Application US/09445023A
; Patent No. US20020119167A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka

QY 565 CNLEPCPSSASGKSFRREQ 583
Db 368 CNLEDCPDN-NGKTFREEQ 385
RESULT 12
US-10-097-580-1
; Sequence 1, Application US/10097580
; Publication No. US20030032168A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka
; APPLICANT: Inoguchi, Eiichi
; APPLICANT: Hakezaki, Michinori
; APPLICANT: Ishioka, Keiko
; APPLICANT: Ishida, Yukako
; APPLICANT: Matsushima, Kouji
; APPLICANT: Kuno, Kouji
; TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
; FILE REFERENCE: Q57092
; CURRENT APPLICATION NUMBER: US/10/097,580
; CURRENT FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: 09/445,023
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: JP 9-160422
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 1
; LENGTH: 727
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-580-1
Query Match 42.7%; Score 1351; DB 9; Length 727;
Best Local Similarity 63.6%; Pred. No. 1e-102;
Matches 241; Conservative 50; Mismatches 82; Indels 6; Gaps 4;
QY 209 RAKRFVSIPRYVETLVVADESVMKFGADLEHYLLTLLATAARLYRHPSTILNINIVVK 268
Db 9 RKKRFVSSPRVYETMLVADQSMAEFHGSLKHYLLTLFSAARLYKHPSTIRNSVSLVVK 68
QY 269 VLLLRDRDSGPKVTGNAALTLRNFCANQKLNKVSDDKHPEYWDTAIFTRODLGATTC 328
Db 69 ILVHDEQKGPVTSNAALTLRNFCNQKQHNPPSDRDAEHYDTAIFTRODLGSGTCD 128
QY 329 TLGMADVGTMDKPRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVFGKLRANH 388
Db 129 TLGMADVGTVCPSRSCSVIEDDGLQAFTTAHELGHVFNPHDDAKQKASLNGVND 188
QY 389 MMSPTLIQIDRANPWSACSAAITDLDGSHGDCLLDQSPKPSLPEDLPGASVTL 448
Db 189 MMASMLSNLDHSPWSPCSAYWVTFSLDNGHGECLMDKPNPIKLPDLPGLTYDANRQC 248
QY 449 ELAFVGVSKPCP-YMOYCTKLMCTKAKQOMVQTRHFPWADGTCGEGKLCCLKGACVER 507
Db 249 QFTFGESKPCPDAATCTTLCCTGTSGLLVLCQTKHFPWADGTCGEGKWCVSGKCVNK 308
QY 508 HNLNKH---RVDSWAKWDPYPCSRCTCGGGVQLARRQCTNPTPANGKYCEGVVRKYRS 564
Db 309 TDM-KHFATPVHSGWGPWPGWDCSRTCGGGVQVYTMRECDNPVKNKGKCEGKRVYRS 367
QY 565 CNLEPCPSSASGKSFRREQ 583
Db 368 CNLEDCPDN-NGKTFREEQ 385
RESULT 13
US-09-445-023A-1
; Sequence 1, Application US/09445023A
; Patent No. US20020119167A1
; GENERAL INFORMATION:
; APPLICANT: Hirose, Kunitaka

APPLICANT: Inouchi, Eiichi
APPLICANT: Hakozaiki, Michinori
APPLICANT: Ishioka, Keiko
APPLICANT: Ishida, Yukako
APPLICANT: Matsushima, Kouji
APPLICANT: Kuno, Kouji
TITLE OF INVENTION: Human ADAMTS-1 protein, gene encoding the same, pharmaceutical
FILE OF INVENTION: composition and method of immunologically analyzing human ADAMTS
FILE REFERENCE: Q57092
CURRENT APPLICATION NUMBER: US/09/445,023A
PRIORITY FILING DATE: 1999-12-03
PRIOR APPLICATION NUMBER: JP 9-160422
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 727
TYPE: PRT
ORGANISM: Homo sapiens
US-09-445-023A-1

Query Match 42.7%; Score 1351; DB 10; Length 727;
Best Local Similarity 63.6%; Pred. No. 1e-102;
Matches 241; Conservative 50; Mismatches 82; Indels 6; Gaps 4;
QY 209 RAKREVSIPRVYETLVVADSMVKFPGADLEHLLTLLAARLYRHPSILNINIVVVK 268
Db 9 RKKREVSIPRVYETLVVADSMVKFPGADLEHLLTLLAARLYRHPSILNINIVVVK 68
QY 269 VLLLRDRSGPKVTCNAALTLLNCAWOKLNVSKHPEYDWTAILFTRODLGCATCD 328
Db 69 LLVHDEKQGEVTSNAALTLLNCAWOKLNVSKHPEYDWTAILFTRODLGCATCD 128
QY 329 TLGMADVGTCDPSSCSVIEDDGLPSAFTTAHELGHVFNPNHNDNVKCEVFCKLRANH 388
Db 129 TLGMADVGTCDPSSCSVIEDDGLPSAFTTAHELGHVFNPNHNDNVKCEVFCKLRANH 188
QY 389 WMSPTLIQIDRANPWSACSAAITDLDLSDHGLDLPSPKIPSLDLPDGLASVTLSSQC 448
Db 189 WMSPTLIQIDRANPWSACSAAITDLDLSDHGLDLPSPKIPSLDLPDGLASVTLSSQC 248
QY 449 ELAFGVGSKPCP-YMYCTKLCWCTGKAGQWVCTRHFPWADGTCGEGKCLKAGACVER 507
Db 249 QFTFGEDSKHCPDASTCSTLWCTGTSGVLYVCTKHFPWADGTCGEGKCLKAGACVER 308
QY 508 HNLNKH---RVDSWAKWDYPCSRCTCGGVQVLRARQCTNPTPANGKCYCEGVRYKRS 564
Db 309 TD-RKHEDTPTFGSWGPMGPGWDCSRTCGGVQVLRARQCTNPTPANGKCYCEGVRYKRS 367
QY 565 CNLEPCPSSASGSFREQ 583
Db 368 CNLEDCPDN-NGKTFREQ 385

RESULT 14
US-10-174-590-352
Sequence 352, Application US/10174590
Publication No. US20030008352A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C42

CURRENT APPLICATION NUMBER: US/10/174,590
CURRENT FILING DATE: 2002-06-18
Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 352
LENGTH: 837
TYPE: PRT
ORGANISM: Homo Sapien
US-10-174-590-352
Query Match 42.3%; Score 1337.5; DB 9; Length 837;
Best Local Similarity 45.1%; Pred. No. 1.6e-101;
Matches 265; Conservative 88; Mismatches 192; Indels 43; Gaps 12;
QY 1 MLLGLILTLAFAGRTAGGFEPEVVPVIRLDPDINGRRYRWGPEDSGDGLIFITAF 60
Db 37 LLLLLLTLASLLPSARLASPLREEIIVFPEKLSVIL-----PGSGAPARLLCLQAF 88
QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVLPLQGLTGGSDLRRCFYSGDVNAEPDSFAAVSLC 120
Db 89 GETLLELEODSGVQVEGLTVQYLGOAPE-LLGGAEP--CTYLTGTINGDPESVASLHWD 145
QY 121 GG-LRGAFGYRGAEYVISPINASAPAAQRNSQGAHLQRRGVPGSPGDPTRSCGVASG 179
Db 146 GGALLGVLYQVGAELHLQPLEGGTPNSA--GGPGAHLIRK-----SPASQGGPMCV--- 196
QY 180 WNPAILRALDPYKPRAGFCESRRSRSGRAKRFVSIPIRYVETLVVADSMVKFPGADLE 239
Db 197 -----KAPLGSPSPRPR--RAKRFASLSRFVETLVVADSMVKFPGADLE 239
QY 240 HYLLTLAATLAARLYRHPSILNINIVVVKVLLLRDRSDGPKVTGNAALTLRNFCAMQKKL 299
Db 240 RYLLTYMAAAKAKFHPISIRNPVSVLVVTRVLVILGSGEGPQVGPSPAAQTLRSECAWQGL 299
QY 300 NKVSDKHPEYDWTAILFTRODLGCATCDTLMADVGTCDPKRSCSVIEDDGLPSAFTT 359
Db 300 NTPEDSGPDHFDTAIFLTRODLGCATCDTLMADVGTCDPKRSCSVIEDDGLPSAFTT 359
QY 360 AHELGHVFNPNHNDNVKCEVFCKL-RANHMSPTLIQIDRANPWSACSAAITDLDLSDG 418
Db 360 AHELGHVFNPNHNDNVKCEVFCKL-RANHMSPTLIQIDRANPWSACSAAITDLDLSDG 418
QY 419 HGCLLDLQSPKIPSLDLPDGLASVTLSSQCELAFGVSKPCPYM-OYCTKLMTCTGKAKQ 477
Db 420 YGHCLLDLQSPKIPSLDLPDGLASVTLSSQCELAFGVSKPCPYM-OYCTKLMTCTGKAKQ 477
QY 478 MVTOTRHPWADGTCGEGKCLKAGACVERHNLNKHVD--GSWAKWDYPCSRCTCGGG 535
Db 480 AMQTHSPWADGTCGEGKCLKAGACVERHNLNKHVD--GSWAKWDYPCSRCTCGGG 535
QY 536 VQLARQCTNPTPANGKCYCEGVRYKRSCLNLEPCPSSASGSFREQ 583
Db 540 VQFSSRDCITPVRNGKCYCEGRRTRFRSCNTEDCP-TGSALTFFREQ 586

RESULT 15
US-10-176-758-352
Sequence 352, Application US/10176758
Publication No. US20030008353A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C42

```
; FILE REFERENCE: P3430RIC104
; CURRENT APPLICATION NUMBER: US/10/176,758
; CURRENT FILING DATE: 2002-06-21
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 352
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-352

Query Match      42.3%; Score 1337.5; DB 9; Length 837;
Best Local Similarity 45.1%; Pred. NO. 1.6e-101;
Matches 265; Conservative 88; Mismatches 192; Indels 43; Gaps 12;

Qy 1 MLLGLILTLAFAGTGGFEPEVGVVPIRLDPDINGRRYYWRGPEDSGDQGLIFOITAF 60
Db 37 LLLLLLASLLPSARLASPLREEIIVFEKLVSVL-----PCSGAPARLLCRQAF 88
Qy 61 QEDFYHLTPDAQFLAPAFSTEHLGVLQGLTGGSSDLRRCFYSGDVNAEPDSFAAVSLC 120
Db 89 GETLLLEQDSGVQVVEGLVQVILQGAPE--LLGGAEP--GTyltGTINGDPESVASLHWD 145
Qy 121 GG-LRGAFGRGAYVTSPLPNASAPAAQRNSOGAHLQRRGVPGGSPGDPISRCGVASG 179
Db 146 GGALLGVLYQYRGAEHLQPLEGGTPNSA--GGPCAHLRRK----SPASQGGPMCNV--- 196
Qy 180 WNPAILRALDPYKRRRAGFESRRSRGRKRFVSIPIRYVETLVVADESVMVKFHGADLE 239
Db 197 -----KAPLGSPRPR--RAKRFASLSRFVETLVVADDKMAAFHGAGLK 239
Qy 240 HYLTLTLAARLYRHPISILNPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCAMQKKL 299
Db 240 RYLLTVMAAAKAFKHPISIRNPVSLVTVRLVILSGEGEPQVGPAAQTLRSFCAMQRL 299
Qy 300 NKVSDKHPEYWDTAILFTRODLGATTCTDLGMADVGTMCDPKRSVIEDDGLPSAFTT 359
Db 300 NTPEDSGPDHFDTAILFTRODLGAVSTCDTLGMADVGTVCDPARSCAIVEDDDGLQSAFTA 359
Qy 360 AHELGHVFNPHDNVVKYCEEVFGKL--RANHMSPTLIQIDRANPWSACSAAIITDFLDG 418
Db 360 AHELGHVFNMLHDSKPCISLNGPLSTRHVMAFVMAHVDFEFPWFC SARFITDFLDNG 419
Qy 419 HGDCLLDQPKPISLPEDLPASVYLSQQCELAFGVSGKPCPYM-QYCTKLWCTGKAKGQ 477
Db 420 YGCLLDKPEAPLHPVTFPGKDYADROQLTFGPDHRHCQPLPPPPCAALWCSGHLNGH 479
Qy 478 MVCOTRHPWADGTSCEGKLCCLKGACVERHNLNKHVD--GSWAKWDYPGCSRTC GGG 535
Db 480 AMCOTKHPWADGTPCGPAQACMGGRCLHMDQLQDFNIPAGGWPMPGWDGCSRTC GGG 539
Qy 536 VQLARRQCTNPTPANGKYCEGVKRYRSCNLEPCPSSASGKSFREQ 583
Db 540 VQFSRDCTRVPVNGKYGCEGRTRFRSCNTEDCP-TGSALTFRREQ 586
```

Search completed: April 29, 2003, 17:21:53
Job time : 16.9594 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 13 4848 Seconds
(without alignments)
4156.253 Million cell updates/sec

Title: US-10-009-332-1_COPY_1_583
Perfect score: 3163
Sequence: 1 MLLGLTLTAFAGTAGGFE.....SCNLEPCPSASGKSFREQ 593

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1562.5	49.4	951	2 T00017	gene ADAMTS-1 prot
2	1336.5	42.3	837	2 T00355	hypothetical prote
3	941.5	29.8	2165	2 T21371	hypothetical prote
4	775	24.5	1205	2 T18517	procollagen N-endo
5	684	21.6	550	2 T47158	hypothetical prote
6	433	13.7	1444	2 T18856	angiogenesis inhib
7	393	12.4	860	2 T16892	hypothetical prote
8	329	10.6	957	2 T15976	hypothetical prote
9	329	10.4	903	2 S60257	meltrin alpha - mo
10	309	9.8	617	2 S48160	metalloproteinase
11	307	9.7	411	1 HYSNFA	fibrolase (EC 3.4.
12	306	9.7	571	2 S24789	jararhagin C precu
13	304	9.6	407	2 S66260	metalloproteinase
14	302.5	9.6	609	2 S55270	catalocollastatin p
15	301	9.5	826	2 A60385	monocyte surface a
16	297	9.4	480	1 A30065	trigamin precursor
17	297	9.4	549	2 S48169	metalloproteinase
18	292	9.2	478	2 JC4880	metalloproteinase
19	290.5	9.2	481	2 JC4342	fibrinolytic metal
20	289	9.1	414	2 S41609	atrolysin C (EC 3.
21	284	9.0	414	1 HYSRAC	atrolysin C (EC 3.
22	281	8.9	789	2 S28259	androgen-regulated
23	280	8.9	481	2 S43125	triamucin precursor
24	273.5	8.6	478	2 A43296	atrolysin E (EC 3.
25	268	8.5	616	2 A55796	ecarin precursor -
26	265	8.4	414	2 S41608	atrolysin B (EC 3.
27	255	8.1	610	2 JC7530	vascular apoptosis
28	255	8.1	814	2 G02390	disintegrin-like m
29	243.5	7.7	776	2 S28258	androgen-regulated

30	243	7.7	478	2 JQ1301	hemorrhagic protei
31	231.5	7.3	508	2 T22836	hypothetical prote
32	229.5	7.3	1170	1 TSHUP1	thrombospondin 1 p
33	224.5	7.1	1170	2 A40558	thrombospondin 1 p
34	218.5	6.9	419	2 S41607	atrolysin A (EC 3.
35	217.5	6.9	419	2 A59414	metalloproteinase
36	215.5	6.8	825	2 S55060	feritin alpha-II
37	215	6.8	670	2 T65967	disintegrin-like m
38	213.5	6.7	1042	2 T26644	hypothetical protei
39	210	6.6	416	2 A37877	hemorrhagic protei
40	210	6.6	524	2 S38539	disintegrin-like m
41	202.5	6.4	823	2 S18968	cyritestin precurs
42	200.5	6.3	756	2 S47656	TMDC II protein -
43	200	6.3	952	2 T18900	disintegrin and me
44	197	6.2	1074	2 JC5928	semaphorin F precu
45	196.5	6.2	905	2 S55059	feritin alpha-I -

ALIGNMENTS

RESULT 1

T00017

gene ADAMTS-1 protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000

C:Accession: T00017

R;Kuno, K.; Lizaso, H.; Ohno, S.; Matsushima, K.

Genomics 46, 466-471, 1997

A:Title: The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1

A:Reference number: Z14055; MUID:98110583; PMID:9441751

A:Accession: T00017

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-951 <KUN>

A:Cross-references: EMBL:AB001735; NID:g2809056; PIDN:BAA24501.1; PID:g2809057

A:Superfamily: thrombospondin type 1 repeat homology

F:542-598/Domain: thrombospondin type 1 repeat homology <THR3>

C:Gene: ADAMTS-1

A:Introns: 228/1; 343/3; 388/1; 444/1; 539/3; 602/1; 660/3; 719/2

Query Match 49.4%; Score 1562.5; DB 2; Length 951;

Best Local Similarity 50.4%; Pred. No. 7.4e-110;

Matches 315; Conservative 79; Mismatches 154; Indels 77; Gaps 16;

QY 1 MLLGLTLTAFAGTAGG--FEPEREVVPILRDPDINGRRYYWRGP-EDSGDQGLIFOI 57

Db 20 LLLASITMLCARGAHRPTEDDELVP-SLE-----RAPGHDSTTRL--RL 66

QY 58 TAFQEDFYHLPTDPAQFLAPAFSTEHLGVPLQGLTGGG-----SDLRRFCYSGDVNA 109

Db 67 DAFQOQLHLKLPDPSGLFAPGFTLTQV-----GRSPSEAQLDPTGDLAHCFYSGTVNG 121

QY 110 EPDSEFAVSLCGLRCAGFYRGAEVVISPLPNAS----APAAQRNSQGA----HLLQRRG 161

Db 122 DFGSAALSLCGVRCAGFYLGQEEFFIQAPGVATERLAPAVPEESSARPOFHILRRR- 180

QY 162 VFGPGSGDPTSCGVASGWNPAIRALDPYKPRRAGFSGESRRRR-----DETLDTSRSPESQNRQWVRDTPDQACKP 206

Db 181 ----RRSGGAKCGVMD-----DETLDTSRSPESQNRQWVRDTPDQACKP 226

QY 207 ----SGRAKRFVSIPIRYVETLVVADESVMKFGADLEHYLLPLLTAAARLYRHPHSILNPI 262

Db 227 SGPGSIRKKRFVSSPRYVETMLVADQSMADFHGSLKHVLLTFLSVAARFYKHPHSIRSI 286

QY 263 NTWVKVLLLRDSDGPKVTGNAALTNRNFCQWKLKNSDKHPHYWDLTALLFTRODLIC 322

Db 287 SLVWVKILVIYEQKGPEVTSNAALTNRNFCQWKLKNSDKHPHYWDLTALLFTRODLIC 346

QY 323 GATTCDTLGMADVGTMCDPKRCSCVIEDDGLPSAFTTAHELGHVFNMPHDNVKVEEVEFG 382

Db 347 GSHTDCTLGMADVGTCDPSPRSVIEDDGLQAFTTAHELGHVFNMPHDDAKHCASLNG 406
QY 383 KLANHMSPTLIQIDRANPWSACSAIITDFLDGSHGCLLDOPSKPISLPEDLPAGSY 442
Db 407 VTGSHLMASLSSLDHSPQSPSCSAIYMTVTFDNGHGHGCLMDKPNPKLPSDLPLCTLY 466
QY 443 TLSQOCFLAFGVGSKPCP-YMOYCTKLWCTGKAGQWQVQTRHPFWADTSCGEGKLCUK 501
Db 467 DANRQCOFTGESKHCPCDAASCTTLMCTGTSGLLVCQTKHPFWADTSCGEGKWCYS 526
QY 502 GACVERHNLNKH---RVDSGNKWDYGPSCSTCGGQVQLARRQCTNPTNPANGKGYCEGY 558
Db 527 GKCYNKTDK-KHEATPVHGSWGPWGDCSRTCGGQVQYTMRECDNPVPKNGKYCEGK 585
QY 559 RVYRSNLEPCPSSASGASGAFREEQ 583
Db 586 RVYRSNLEPCPSSASGASGAFREEQ 609

RESULT 2
T00355
hypothetical protein KIAA0688 - human
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00355
R:Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5, 169-176, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A:Reference number: Z14142; MUID:98403880; PMD:9734811
A:Accession: T00355
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-837 <I>SH>
A:Cross-references: EMBL:AB014588; NID:G3327189; PIDN:BAA31663.1; PID:G3327190
A:Experimental source: Brain
C:Genetics:
A:Gene: KIAA0688
C:Superfamily: thrombospondin type 1 repeat homology
F:519-575/Domain: thrombospondin type 1 repeat homology <THR3>

Query Match 42.3%; Score 1336.5; DB 2; Length 837;
Best Local Similarity 45.1%; Pred. No. 7.3e-93;
Matches 265; Conservative 88; Mismatches 192; Indels 43; Gaps 12;

QY 1 MLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYWRGPDGSLIFQITAF 60
Db 37 LLLLLLASLLPSARLASPLREEIVPEKLNGSVL-----PGSGTPTARLLICRLQAF 88

QY 61 QEDFYHLTPDAQFLAPAFSTEHLGVPLGLTGSSDLRCFYSVDYNAEPDFAAVSLC 120
Db 89 GETLLELDQSGVOVEGLTVQYLQQAPE-LLGGAEP-GTYLGTINGDPESVASLHWD 145

QY 121 GG-LRGAFYRGAEYVISPLNAPAAQRNSOQAHLLQRRGVPGPSPGDPSTRCGVASG 179
Db 146 GGALLGVLYRGAEHLQPLGGTPNSA--GGPGAHILRRK-----SPASGGQPMCNV--- 196

QY 180 WNPAILRALDYPKPRAGGSGRRSGRAKRFVSPRYVETLVVADESVMKFGADLE 239
Db 197 -----KAPLGSPSPRPR--RAKRFASLSRFVETLVVAADDKMAAFPHGAGLK 239

QY 240 HYLLTLLATAARLVHRHPSILNIPINIVVVKVLLLRDRDSGPKVTGNAALTLRNFCWAKKL 299
Db 240 RYLLTVMAAAAKAFKHPISRPVSLVTVRLVILSGEGEPQVGSAAQTLSFCWARGL 299

QY 300 NKVSDKHPYWDTAITFRQDLGCATCTDGLMADVGTMDCKPKRSCSVIEDDGLPSAFTT 359
Db 300 NTPEDSDPHEDTAITFRQDLGVSTCDTLGMADVGTVCVPARSCAIVEDDGLQSAFTA 359

QY 360 AHELGHVFNMPHNDNVKVEEFGKL-RANHMSPTLIQIDRANPWSACSAIITDFLDG 418
Db 360 AHELGHVFNMLHNSKPCISLNGPLSTSRHVNAPYMAHVDPPEEPWSPCSARFITDFLDNG 419

QY 419 HGDCLLDQSPKISLPEDLPAGSYTLSQOCFLAFGVGSKGPCPYM-QYCTKLWCTGKAGQ 477

Db 420 YGHCLLDRPEAPLHLPTVFPKDYADRQCQLTFGPDSSRHCPQLPPPCALWCSGHLNGH 479
QY 478 MVCOTRHPFWADGTSCGEGKLCGLACACVERHNLNKHRYD--GSWAKWDYGPSCSRTCGG 535
Db 480 AMCQTKHSFWADGTPCGPAQACMGRCGLHMDQLDFNIPQAGGWPWGPCDSCRTCGG 539
QY 536 VOLARRQCTNPTNPANGKGYCGVRYKRSNLEPCPSSASGASGAFREEQ 583
Db 540 VQFSSRDCTRPVPRNGKYCEGRRTRFRSCNTEDCP-TGSALTFRREQ 586

RESULT 3
T21371
hypothetical protein F25H8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T21371; T24896
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19413
A:Accession: T21371
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WIL>
A:Cross-references: EMBL:Z69360; PIDN:CAA93287.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone F25H8
R:Gajadaty, S.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z19949
A:Accession: T24896
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2165 <WIL>
A:Cross-references: EMBL:Z69361; PIDN:CAA93288.1; GSPDB:GN00022; CESP:F25H8.3
A:Experimental source: clone T13H10
C:Genetics:
A:Gene: CESP:F25H8.3
A:Map position: 4
A:Introns: 31/1; 52/1; 135/2; 193/3; 216/1; 266/1; 495/2; 547/3; 584/3; 634/2; 744/1;

Query Match 29.8%; Score 941.5; DB 2; Length 2165;
Best Local Similarity 38.7%; Pred. No. 1.6e-62;
Matches 203; Conservative 75; Mismatches 181; Indels 65; Gaps 14;

QY 101 CFYSGVDNAPDSFAAVSLC---GGLRGAFYRGAEYVISPLNPSA---PAAQRNSOGA 154
Db 169 CIYRAHVKG-VHQHSTVNLCDSEDLGYLMLALPSGIHTVETPIISGNGTEHDGASHRQ--- 225

QY 155 HLLQRRGVPGSGDPTSRGV-----ASGNPAILRALDYPKPRRAGFSGRSR 204
Db 226 HLVRKFDPMHFKSFDHLNSTSVNETTETTVATWQDQWEDVIER-----KARSR 272

QY 205 RRSGRAKRFVSPRYVETLVVADESVMKFGADLEHYLLTLLATAARLVHRHPSILNIPIN 264
Db 273 RAAN-----SWDHYVEVLVWADTKMYEYHGRSLEDYVLTFTVASIYRHOSLRASIN 326

QY 265 VVYKVVLLLRDRDSGPKVTGNAALTLRNFCWAKKLNVKSDKHPEYWDTAITFRQDLCA 324
Db 327 VVYKLVLTENAGPRITQNAOQTLODFCRQWQYYNDPDDSSVQHHIDVAILLTRDKICRS 386

QY 325 T-TCDTLGMADVGTMDCKPKRSCSVIEDDGLPSAFTTAAHELGHVFNMPHNDNVKVEE 380
Db 387 QGKCDTLGLAELTMDCMQKSCAIIEDNGLSAAFTIAHELGHVFSIPHDDERKCSYMPV 446

QY 381 -----FGKLRAN---HMSPTLIQIDRANPWSACSAIITDFLDGSHG--DCLLD 425
Db 447 NKVCKFGSTKFDKTFQNNFHMIAPTLEYNTHPWSWSPCSAGMLERFLENNGQTCLDF 506

QY 426 QPSKPSISLPE---DLPGASYTLSQOCFLAFGVGSKGPCPYMOYCTKLWCTGKAGQWQV 481
Db 507 QPERRYIEDVFRDEPGKDYDAHQCKQFVGFGPAGSELCPYMPYTCRLWCATFVGSOMGR 566

Db 524 KGPLDGTMCAPGRKHCFCKGHCILWLPDILKR-----DGNWGAWSFGSGSCFTCTGTV 575
QY 537 QLARRQNTTPTANGPKYCEGVVRKYRSCNLPFCPSASGKSFREQ 583
Db 576 KFRTRQCDNPHANGGRTCSGLAYDFQLCNSQDCPDALA--DFREQ 620

RESULT 5
T47158
hypothetical protein DKFzp762C1110.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47158
R:Blum, H.; Bauersachs, S.; Kewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24379
A:Accession: T47158
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-550 <AAA>
A:Cross-references: EMBL:AL162080
A:Experimental source: adult melanoma (MeWo cell line); clone DKFzp762C1110
C:Genetics:
A:Note: DKFzp762C1110.1

Query Match 21.6%; Score 684; DB 2; Length 550;
Best Local Similarity 59.2%; Pred. No. 8.1e-44;
Matches 119; Conservative 25; Mismatches 51; Indels 6; Gaps

QY 387 NHMSPFLIQIDRANPWSACSAIIITFDLSGHGDCLLDQPSKPTSLPEDLPASGYTLISQ 446
Db 10 SHMASMLSLNDH5QSPSCSAYMITSLFDNGHGECMLDKPQNPILQPLDGLPSTSYDANR 69
QY 447 QCELAFGVGSKPCP-YMQYCTCKLWCTGKAKGMWCVOTRFPWADGTSCEGKLCCLKGACV 505
Db 70 QCQTFGDSKHCPCDAASTCTLWCTGSGVLVCOTKHFPAWADGTSCEGKWCINGKCV 129
QY 506 ERHNLNKH---RVDGWSKAWDPYGPCSRTCCGGVQLARRQCTNPTPANGKYCEGVVRKY 562
Db 130 NKTD-RKHFDTPFHGSGWGMGMPGWCDSRTCCGGVQVYTMRECDNPVPKNGKYCEGKRVRY 188
QY 563 RSCNLEPCPSSASGKSFREQ 583
Db 189 RSCNLEPCPDN-NGKTFREQ 208

RESULT 6
T18856
angiogenesis inhibitor homolog - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T18856; T24653
R:McMurray, A.
submitted to the EMBL Data Library, July 1995
A:Reference number: Z19031
A:Accession: T18856
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1444 <WIL>
A:Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN00028; CESP:C02B4.1
A:Experimental source: clone C02B4
R:McMurray, A.
submitted to the EMBL Data Library, July 1995
A:Reference number: Z19917
A:Accession: T24653
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1444 <W12>
A:Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:C02B4.1
A:Experimental source: clone T07C5
C:Genetics:
A:Gene: CESP:C02B4.1

A:Map position: X
A:Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3; 568/3

Query Match 13.7%; Score 433; DB 2; Length 1444;
Best Local Similarity 25.9%; Pred. No. 2.3e-24;
Matches 156; Conservative 81; Mismatches 264; Indels 102; Gaps 24;

QY 43 RGPEDSGDGLIQTAFQEDFYHLTPDQAFLAP-AFSTEHLGVLPQGLTGGSSDLRRC 101
DB 36 RHLPLKGN--LKKMFTAWNDYHLNLKRNKRISVPHIISVVRHGDDDDVTYAGLRDYEQC 93
QY 102 FYSGDVNAEPDSFAAASLCGLRGAFGYRGAËYISPLPNASAPAAQNRNSQGAHLQRRG 161
DB 94 HYQGEVKSNGMKRAAISDCGALMGSIWMEHFLVLQLPK----RVHHLQKRLVYKRS 149
QY 162 VPGPGSDPTSRGCVASGWNPAITRALDPYKPPRAGFGESESRSSRRSRGRKRFYSIPRY-- 219
DB 150 -----AGLLTNAESKTRREETRLOEQESFCDT-SEQLDDPA---WTIPAHLLH 193
QY 220 ----VETLVVADESQWKFHGAD-----LEH-----YLTLL 246
DB 194 FNVTIPTSAQLDSSFI-FPNMNDPITLIGFLDSKLFEPEREYIQDAEQHLLFSLALI 252
QY 247 ATAARLYRHPISILNPINIVVVKVLLLRDRDS---GPKVTGNAALTLRNFCAMQKLNKV 302
DB 253 NNHVLYQQDTLPNLDIVIVVEMWTQPSALSSTGVHKNQQAQSLDAFCRYQAHNPG 312
QY 303 SD-KHPYWDYTAFLTTRQDLCGGATCTDTLGMADVGMCDPKRSCSVTEDDGLPSAFTAH 361
DB 313 TDLTDMNHYDHGVLTLTYDIDYHTTT-SVAGVAPVARMCDPLFACSLVEGLHGRSFVIAH 371
QY 362 ELGHVFNPNHNDVK-VCEEVEFGKLRANHMSPITLIQIDRANPWSACSAIITDPL----D 416
DB 372 EMGHNMGVMHDGVQNCNKGCLMSAVNGAKT-----TWSDCSVREENAFLLQLDE 423
QY 417 SGHGDCLLDQSPKISLPE-----DLPCASTLSQCELAFCGVGSK-PCP-----YMQYCTK 467
DB 424 SGRCNCLRDASPLGISTNNHLSURLPGORFTADQOCYSFWGRDYKYVEIPNGKAMDDICRI 483
QY 468 LWCMTGKAGQVMVQTRHFPWADGTSCEGKGLCLKGCACVE-RHNLNKRHVRDVGSAKWPD--- 523
DB 484 LWCNNGSGS---TISTAH-PALEGSWCAGANKWCHKQCTHTFGLTPVPIDGENSEWGAE 539
QY 524 ---PYGCSRTCGGGVQOLARQCTNPTTPANGKVCYGVRYKRSCLNEPCPSSASGKSFR 580
DB 540 KGCPIDQCAVSGSITVQGGHQHRCVNPANPNNGKTKCEGANINGIVCG-----ATSSNCLGFT 595
QY 581 EEQ 583
DB 596 REE 598

RESULT 7
Tl6892
hypoetical protein Tl9D2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: Tl6892
R:Bentley, D.
submitted to the EMBL Data Library, December 1995
A:Description: The sequence of C. elegans cosmid Tl9D2.
A:Reference number: Z18599
A:Accession: Tl6892
A:Status: preliminary;
A:Molecule type: DNA
A:Residues: 1-860 <BEN>
A:Cross-references: EMBL:U42846; NID:g1125809; PID:g1125810; PIDN:AAA83600.1; CESP:Tl9D2.1
A:Gene: CESP:Tl9D2.1
A:Introns: 56/3; 96/1; 171/3; 199/3; 245/2; 293/2; 335/3; 395/3; 426/3; 484/2; 505/3; 568/3

Query Match 12.4%; Score 393; DB 2; Length 860;
Best Local Similarity 29.8%; Pred. No. 1.3e-21;

Db 276 IRYEVLTRPSALAGYLNHNGNAOMYLDRCYQORNL-AVRD-----WDHAIMLTGYDIH 329
QY 322 CGATTCTDLGMADVTCMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEVF 381
Db 330 RGAGSRISISGLARLDGCMCPWNTCTLABGLDFTSAFIGHTHELGHFRFTLKD 381
QY 382 KGLRANHMSPYL-----IQIDRANPWSACS-----AALIT-----DFLDSGHG 420
Db 382 ---KSDTLFTGTFCGSPKWCQGLRCVPTWGTNEIQTPVQHVAPVTVTLPSRIDGSWSGW 438
QY 421 DCLLDQ-----PSKIPISLPEDLPAS-----YTLSSQCELA----- 451
Db 439 ATICSQCCTNGILSGVGLAIARRTCSAPYPANGGSDCVGSTSRAVLCSROCGRAKSWDE 498
QY 452 -----FGVSRKCPVMQYCTKLWCTGKAKGOMVQCTRHFPW-----A 498
Db 499 YISDKCMEQRLKNDRELTKGSGOLNRPFPQACKVFCDVQ---OHYGSORNYRFFGDNLP 555
QY 489 DGTSGGEGKLCGACVERHNLKHRVD-----GSAKWDPYPG 527
Db 556 DGTSGVDYRLDCEGLALNCLNALLISRDQSCPTDTCPTDQSSSVYRGOWGTWSLWTS 615
QY 528 CSRTCGGVQLARQCINPTPANGKVCCEGVYKYSCLNLEPCPS 572
Db 616 CTATCGGYRKNRACS-----ITG--QCEGNEDETEVCSSCPS 654

RESULT 9
S60257
metlin alpha - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 20-Jun-2000
C:Accession: S60257
R:Yagami-Hiromasa, T.; Sato, T.; Kurisaki, T.; Kamiyo, K.; Nabeshima, Y.I.; Fujisawa-Seh
Nature 377, 652-656, 1995
A:Title: A metalloprotease-disintegrin participating in myoblast fusion.
A:Reference number: S60257; MUID:96026308; PMID:7566181
A:Accession: S60257
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-903 <VAG>
A:Cross-references: EMBL:D50411; NID:g1054586; PIDN:BA08912.1; PID:g1054587
C:Superfamily: mouse metlin alpha; disintegrin homology
F:421-503/Domain: disintegrin homology <DIS>
F:349/Active site: Glu #status predicted

Query Match 10.4%; Score 329; DB 2; Length 903;
Best Local Similarity 21.5%; Pred. No. 9.4e-17;
Matches 159; Conservative 74; Mismatches 245; Indels 262; Gaps 27;

QY 2 LLLGILFLAFAGRTA-----GGFEPEYVVPVIRLDPDINGRRYWRGPDSDGQ 52
Db 15 LLLALAGALLAPRAAGMSLWDQRGAYEVARASL--LSKDPGIPGQSI---PAKDHPDV 68

QY 53 LIQITAFQEDFYLHLTPDAQFLAPAFSTHGLVPLQ-----GLTGGSDLRRCFFVSGDV 107
Db 69 LTVOLQLESRLILSLERNEGLANGFTETHY---LDQGTDSLTRNHTD--HCYVGHV 123

QY 108 NAEFDSFAAYSLCGGLRGAFYRGAEEYVISPLN-----APA-AORNSQGAHLRLQRRG 161
Db 124 QGDAASVSLSTCSDLRGLIMEFNKYSLPEMKNTTDSYKLVPAESKTIQGL----- 176

QY 162 VPGPSPDPTSRGCVAGSNPAILRALDPYKPRAGGSRSSRRSRRAKRFVSIPIRYVE 221
Db 177 -----CGSHNKSNTMEDVSP-----GTSQMRARRHK---RETLMKTKYVE 215

QY 222 TLVVADESMVKFPGADLEHYLLTLLATAARLYRHPISILNPINLVVYKVLILLDRDSGPVK 281
Db 216 LVIVADNREFORQKLEKVKQRLIEANHVDFYRPLN-IRLVGVGVWNDIDK-CSEI 273

QY 282 TGNALTLNRNFCAWQKLVKSDKHPYWDTAILFTRODLGATTCDTLGMDADVGTMCDDP 341
Db 274 SQDPFTRLHFEFLDWRIKLLPRKSH---DNAQLISGVYFQGT---TIGMAPIMSCTA 325

QY 342 KRSCSVI---EDDGLPSAFTTAHELGHVFNPHDNV-KVCEEVFGKLRANHMSPYLIQI 397
Db 326 EQSGVGVMDHSDSPFLGAAYVTLAHELGHNFNMNHDTLERGCSCRMAAEKGCCIMNPS----- 381
QY 398 DRANP---WSACSAALITDFLDSGHGDCLLDOP----- 427
Db 382 -TGFPFPMVFFSSCRKDLASLEKGMGMCLFNLPEVKAQFAGKCGKNGYVEEGEEDCGE 440
QY 428 -----CTKLWCTGKAK-----COMVCQTRHFPW 487
Db 441 PEECTNRCCNATTCTLKPDVAHAGOCCEQCLKPPGTACTACRGSSNSCDLPFECTGTAPHC 500
QY 442 -----YTLSSQCELAFCVGSKPCPYMOY----- 464
Db 501 PANVYLHDGHPGQVDGYCYNGICQTHQOCVTLWPGGAKPAPGICFERVNSAGDPYGC 560
QY 465 -----CTKLWCTGKAK-----COMVCQTRHFPW 487
Db 561 GRDSKSAFAKCELRDAKCKICQCGGASRPVGTNAVSIETNIPQOEGRIILCRGTHVYL 620
QY 488 AD-----GTSCGEGKLCGACVERHNLKHRVDGWSWAKWDPYGPCSRTC-GGGV 536
Db 621 GDDMPDPGLVLGATKCAEGKICLNRRCQNISVFGVHK-----CAMQCHGRGV 667

QY 537 QLARROCTNPTPANGKGYCE 556
Db 668 CNNRKNC-----HCE 677

RESULT 10
S48160
metalloprotease (EC 3.4.24.-) H-I precursor - carpet viper
N:Contains: disintegrin
C:Species: Echis pyramidum leakeyi
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jul-2000
C:Accession: S48160
R:Paine, M.J.I.; Moura-Da-Silva, A.M.; Theakston, R.D.G.; Crampton, J.M.
Eur. J. Biochem. 224, 483-486, 1994
A:Title: Cloning of metalloprotease genes in the carpet viper (Echis pyramidum leakeyi)
A:Reference number: S48160; MUID:95010025; PMID:7925363
A:Accession: S48160
A:Molecule type: mRNA
A:Residues: 1-617 <PAT>
A:Cross-references: GB:X78970; NID:g763092; PIDN:CAA5565.1; PID:g763093
C:Superfamily: mouse metlin alpha; disintegrin homology
C:Keywords: hydrolase; metalloprotease; venom
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-192/Domain: propeptide #status predicted <PRO>
F:193-617/Product: metalloprotease H-I #status predicted <WAT>
F:403-485/Domain: disintegrin homology <DIS>
F:337/Active site: Glu #status predicted

Query Match 9.8%; Score 309; DB 2; Length 617;
Best Local Similarity 25.3%; Pred. No. 1.9e-15;
Matches 119; Conservative 60; Mismatches 206; Indels 86; Gaps 17;

QY 45 PEDSGDGLIQTAFQEDFYLHLTPDAQFLAPAFSTHGLVPLQGLTGGSDLRRCFFYS 104
Db 50 PEQKTEDTRVEYFKVNGEPVVLHLEKNKGLFSEYTHSPDGREITNPNVEDHCYH 109

QY 105 GDVNAEPDSFAAVSLCGGLRGAFYRGAEEYVISPLN-----PNASAPAAQRNSOGAHLRLQRRG 162
Db 110 GRIQNDADSTASMSACNLKGYFMLRGETYLLIEPLKIPDSEAHAVKYENVE----- 161

QY 163 PGGPSPDPTSRGCVAGSNPAILRALDPYKPRAGGSRSSRRSRRAKRFVSIPIRYVE 221
Db 162 ---REDEAPKMGCVTQTNWES-----DELKASOLVATSEQR-----SVKYIE 203

QY 222 TLVVADESMVKFPGADLEHYLLTLLATAARLYRHPISILNPINIVV-----VKVLLLR 273
Db 204 FVVVADYIMYRKNND-----STAVRRRIYEIVNLMNMYIVFNHVALTHIEIWSTR 256

QY 274 DRDGGKVTGNAALTLRNFCAWQKLNKVSQKHPEYWDITAILFTRODLGCAATCTDLGMA 333
 Db 257 DQIT---VQSAADVTLDLFGDWRANKLLTRKH---DNAQFTGTINLNG-----QTLGIA 305
 QY 334 DVGTMCDPKRSCVIED---DGLPSAFTTAHELGHVFNPHDNVY-CEEVFGKLRANHM 389
 Db 306 RMRGMCSPNSVGVQIDYCKNLYLLVAITWAHELGHNLGHDHNGCNCNCPDTSCIMSAYAG 365
 QY 390 MSTFLQIDRANPWSACSAALITDFLDSGHGCDLLDPSKPSLSPEDLPASYT-LSQOC 448
 Db 366 PEPVF-----SFSNCSRDYRSFNDSQSKIDNKLPTKTDIVSPVSGNFTVEVGEEC 418
 QY 449 ELAFGVGSKPCPYMQVCTKLMWCTGKAKGMVQOTRFPWADGTSCEGKLC 499
 Db 419 D-----CGSRTYCRNPCCNA-----TTCKL-----TPGSQCADGEC 450
 RESULT 11
 HYSNFA
 fibronase (EC 3.4.24.-) precursor - southern copperhead
 C:Species: Agkistrodon contortrix contortrix (southern copperhead)
 C:Date: 30-Sep-1992 #sequence_revision 09-Apr-1998 #text_change 18-Jun-1999
 C:Accession: S66259; A41827; S25461; A37303
 R:Selistre de Araujo, H.S.; Ownby, C.L.
 Arch. Biochem. Biophys. 320, 141-148, 1995
 A:Title: Molecular cloning and sequence analysis of cDNAs for metalloproteinases from b
 A:Reference number: S66259; MUID:95314311; PMID:7793974
 A:Accession: S66259
 A:Molecule type: mRNA
 A:Residues: 1-411 <SEL>
 A:CROSS-references: EMBL:U18233; NID:g603214; PIDN:AAC59703.1; PID:g603215
 R:Randolph, A.; Chamberlain, S.H.; Chu, H.L.C.; Retzius, A.D.; Markland Jr., F.S.; Masia
 Protein Sci. 1, 590-600, 1992
 A:Title: Amino acid sequence of fibronase, a direct-acting fibrinolytic enzyme from Agk
 A:Reference number: A37303; MUID:93278288; PMID:1304358
 A:Accession: A41827
 A:Molecule type: protein
 A:Residues: 191-192, 'R', 194-199, 'Q', 201-240, 'Q', 242, 'T', 244-357, 'AAM', 361-381, 'T', 383-39
 A:Note: 379-Glu and 382-Leu were also found
 A:Note: A variant lacking 192-Gln was also found
 R:Guan, A.L.; Retzius, A.D.; Henderson, G.N.; Markland Jr., F.S.
 Arch. Biochem. Biophys. 289, 197-207, 1991
 A:Title: Purification and characterization of a fibrinolytic enzyme from venom of the sc
 A:Reference number: S17490; MUID:91378546; PMID:1898066
 A:Accession: S25461
 A:Molecule type: protein
 A:Residues: 331-341 <GUA>
 C:Genetics:
 C:Superfamily: atrolysin C
 C:Keywords: hydrolase; metalloproteinase; pyroglutamic acid; venom; zinc
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-190/Domain: propeptide #status predicted <PRO>
 F:191-411/product: fibronase #status experimental <MAT>
 F:191/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experime
 F:308-388/disulfide bonds: #status experimental
 F:333,337,343/Binding site: zinc (His) #status predicted
 F:334/Active site: Glu #status predicted
 F:348-355,350-372/disulfide bonds: #status predicted
 Query Match 9.7%; Score 307; DB 1; Length 411;
 Best Local Similarity 26.4%; Pred. No. 1.6e-15;
 Matches 112; Conservative 67; Mismatches 179; Indels 66; Gaps 17;
 QY 24 EWVPIRLDPDINGRYWYRGPDSGDLFIQITAFQEDFYHLHTPDAQFLAPAFSTE 83
 Db 32 EWVYPRKVTTPVRG-----AVQPKYDAMQYEFKVGCEPVVHLERKNGLFSDYSETH 85
 QY 84 LGVPLQGLTSGSLLRCFYSGDVNAEPDSPAAYSLGGLRGAFYRGAEYVISPFPNAS 143
 Db 86 YSPDGREITTYPLVEDHCYHYHRIENDADSTASISACNGLKGHPKLGQEMYLIEPL---- 141
 QY 144 APAAQRNSQGAHLLQR-RGVPGGSGDPTSRGCVASGNWNPAILRALDPYKPRRAGFESR 202

Db 142 ---ELSDSEAHAYKYENVE--KEDEAPKMGCVTONW-----ESYEPIKKAQOLNL 187
 QY 203 SRRESGRAKREVSIPRYVETLWVADESM-VKFGAD--LEHYLLTLIATARLYRHPSIL 259
 Db 188 TPEOQGFPO-----RVELVIVADHRMNTKNGSDSKIRQWVHQIVNTINEIYR----- 236
 QY 260 NPINI--VVVKVLLRRDRSGPKVTGNAALTNRNFCAWQKLNKVSQKHPEYWDITAILFT 317
 Db 237 -PLNIRFALVGLIWSNQDL-ITVTSVSHDTLASFGNWRRETDLLRRQRH-----DNAQLLT 290
 QY 318 RQDLGATTCDTLGMADVGTMCDPKRSCTVEDD---GLPSAFTTAHELGHVFNPHDNV 374
 Db 291 AIDFDG-----DTVGLAYVGGMCQKHKSTGVIQDHSAINLLVALTMAHELGNLGMNDG- 345
 QY 375 KVCEVFGKLRANHMWSPTLTIQIDRANPWSACSAALITDFLDSGHGCDLLDQD-----S 428
 Db 346 NQCH-----CGANSCVMPVSLSDQPSKLFSDCSKKQYTFPLPVNPNQCIILNKPLRTDTAS 400
 QY 429 KPIS 432
 Db 401 TPVS 404
 RESULT 12
 S24789
 Jararhagin C precursor - jararaca (fragment)
 N:Alternate names: single chain botrocetin
 N:Contains: disintegrin-like 28K protein; hemorrhagic proteinase (EC 3.4.24.-)
 C:Species: Bothrops jararaca (jararaca)
 C:Date: 20-Feb-1985 #sequence_revision 29-Aug-1997 #text_change 09-Jun-2000
 C:Accession: S24789; JC2245; A44463; A37958; JC2373
 R:Paine, M.J.I.
 submitted to the EMBL Data Library, August 1992
 A:Reference number: S24789
 A:Accession: S24789
 A:Molecule type: mRNA
 A:Residues: 1-371 <PAI>
 A:CROSS-references: EMBL:X68251; NID:962467; PID:962468
 R:Usami, Y.; Fujimura, Y.; Miura, S.; Shima, H.; Yoshida, E.; Yoshioka, A.; Hirano, K
 Biochem. Biophys. Res. Commun. 201, 331-339, 1994
 A:Title: A 28 kDa-protein with disintegrin-like structure (jararhagin-C) purified fro
 A:Reference number: JC2245; MUID:94256999; PMID:8198592
 A:Accession: JC2245
 A:Molecule type: protein
 A:Residues: 360-571 <USA>
 A:Experimental source: venom
 R:Paine, M.J.; Desmond, H.P.; Theakston, R.D.; Crampton, J.M.
 J. Biol. Chem. 267, 22869-22876, 1992
 A:Title: Purification, cloning, and molecular characterization of a high molecular we
 ily.
 A:Reference number: A44463; MUID:93054601; PMID:1385408
 A:Accession: A44463
 A:Molecule type: mRNA
 A:Residues: 1-23, 'O', 25-92, 'G', 94-131, 'G', 133-169, 'Q', 171-571 <PA>
 A:CROSS-references: GB:X68251
 A:Experimental source: venom gland
 A:Note: sequence inconsistent with the nucleotide translation
 A:Note: sequence extracted from NCBI backbone (NCBIP:118104)
 R:Fujimura, Y.; Titani, K.; Usami, Y.; Suzuki, M.; Oyama, R.; Matsui, T.; Fukui, H.;
 Biochemistry 30, 1957-1964, 1991
 A:Title: Isolation and chemical characterization of two structurally and functionally
 A:Reference number: A37958; MUID:91129280; PMID:1993206
 A:Accession: A37958
 A:Molecule type: protein
 A:Residues: 360-372, 'E', 374-378, 'X', 380-384 <FUJ>
 A:Note: 361-Val was also found
 C:Comment: Inhibits collagen- and ADP-induced platelet aggregation.
 C:Superfamily: mouse meltrin alpha; disintegrin homology
 C:Keywords: hydrolase; metalloproteinase; venom; zinc
 F:360-571/product: jararhagin C #status experimental <MAT>
 F:362-444/Domain: disintegrin homology <DIS>
 F:295,299,305/Binding site: zinc (His) #status predicted

Db	265	NFLSWREQ	---NLQGQPH	---DNNVLITGVDPFGS	---TVGLAKVASALCS	---RHSGAVNOD	315
QY	351	--DELPSAFT	TAHELGHVENMPHD	---NVKVC	---EEVFGKLRANHMS	---PTLIQIDR	399
Db	316	HSKWSIGV	ASTMAHELGNLGM	SHDEIP	PCYCPEPREGGCI	MTESIGSKPPI	---370
QY	400	ANPWSACS	AAITITDLSHG	GDCLDOP	---SKPI	---	SLPDLPG 439
Db	371	---FSRCSK	IDLESFVTKPQT	GCLTNV	PDVNRFRVGGP	VCGNLFEVHEGEO	CDGCTPDQCN 427
QY	440	-----	---ASYTL	SOQCEL	-----	---	AFG 453
Db	428	PCCNATT	CLQVKGAE	CSAGTC	HECHKVPK	AGEVCRLSKDK	CDLEEFCDGRKPTCPEDAFO 487
QY	454	VGSFPCP	-----	---YMQY	CTKLW	-----	CTGKAK
Db	488	QNGT	PCPGYCFD	GSCTPLAQ	CDRLMGP	GARVAADSCY	TFSTPPCGNGRMYSGRINRCG 547
QY	477	QMVQ	QTRHFPP	-----	---	---	ADGTSOGEGLKCLKACVER 507
Db	548	ALYCEGG	GKPLER	SECTFSS	NHGYCHAL	GTGSNTIDFEL	VLQGTKEEGKVCMDGSC---604
QY	508	HNLNKH	RYDVG	SWAKWD	YPGPC	SRTC	GGGVQIARRQC 543
Db	605	ODLRVY	RENG	SACN	NNHGV	CN	-----KREC 631

Search completed: April 29, 2003, 17:13:25
Job time : 17.4848 secs

```

Search completed: April 29, 2003, 17:13:23
Job time : 17.4848 secs

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 7.18023 Seconds
(without alignments)
3367.676 Million cell updates/sec

Title: US-10-009-332-1_COPY_1_583

Perfect score: 3163

Sequence: 1 MLLGLILTLFAGRTAGGFE.....SCNLEPCPSSASGKSFREQ 583

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1562.5	49.4	968	1	ATSL_MOUSE
2	1560.5	49.3	967	1	ATSL_HUMAN
3	1547	48.9	967	1	Q9uql1 rattus norv
4	1362	43.1	890	1	ATSL_HUMAN
5	1338.5	42.3	807	1	ATSL_HUMAN
6	1329.5	42.0	905	1	ATSL_MOUSE
7	1242	39.3	930	1	ATSL_MOUSE
8	1232.5	39.0	1629	1	ATSL_HUMAN
9	1212.5	38.3	930	1	ATSL_HUMAN
10	1168.5	36.9	630	1	ATSL_RAT
11	937.5	29.6	997	1	ATSL_HUMAN
12	891.5	28.2	1077	1	ATSL_HUMAN
13	876.5	27.7	1593	1	ATSL_HUMAN
14	788.5	24.9	1211	1	ATSL_HUMAN
15	775	24.5	1205	1	ATSL_HUMAN
16	752	23.8	1205	1	ATSL_HUMAN
17	725.5	22.9	860	1	ATSL_HUMAN
18	599	18.9	207	1	ATSL_BOVIN
19	599	18.9	245	1	ATSL_BOVIN
20	351	11.1	824	1	AD08_HUMAN
21	348.5	11.0	776	1	AD28_MACFA
22	329	10.4	903	1	AD12_MOUSE
23	326	10.3	813	1	AD33_HUMAN
24	324	10.2	774	1	AD28_MOUSE
25	323.5	10.2	775	1	AD28_HUMAN
26	314	9.9	909	1	AD12_HUMAN
27	306	9.7	571	1	DISJ_BOVJA
28	301	9.5	826	1	AD08_MOUSE
29	300	9.5	956	1	AD19_HUMAN
30	297	9.4	480	1	DISA_TRIGA
31	296	9.4	413	1	ACLA_AGRAC
32	295.5	9.3	920	1	AD19_MOUSE
33	284	9.0	414	1	HRTD_CROAT

ALIGNMENTS

RESULT 1

ID	ATSL_MOUSE	STANDARD;	PRT;	968 AA.
AC	P97857: 054768;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).			
DE	ADAMTS1.			
GN	Mus musculus (Mouse).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OC	NCBI_TaxID=10090;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=129/SvJ;			
PC	MEDLINE=96110583; PubMed=9441751;			
RX	Kuno K., Lizasa H., Ohno S., Matsushima K.; "The exon/intron organization and chromosomal mapping of the mouse ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";			
RA	Genomics 46:466-471(1997).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=97150761; PubMed=8995297;			
RA	Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F., Matsushima K.;			
RT	"Molecular cloning of a gene encoding a new type of metalloproteinase-disintegrin family protein with thrombospondin motifs as an inflammation associated gene.";			
RT	J. Biol. Chem. 272:556-562(1997).			
RL	[3]			
RN	CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.			
RP	MEDLINE=99303657; PubMed=10373500;			
RA	Kuno K., Terashima Y., Matsushima K.;			
RT	"ADAMTS-1 is an active metalloproteinase associated with the extracellular matrix.";			
RT	J. Biol. Chem. 274:18821-18826(1999).			
RL	[4]			
RN	FUNCTION.			
RP	MEDLINE=20389568; PubMed=10930576;			
RA	Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M., Ohno H., Matsushima K.;			
RT	"ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";			
RT	FEBS Lett. 478:241-245(2000).			
RL	[5]			
RN	FUNCTION, AND INDUCTION.			
RP	MEDLINE=20243757; PubMed=10781075;			
RA	Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W., Richards J.S.;			
RT	"Progesterone-regulated genes in the ovulation process: ADAMTS-1 and cathepsin L proteases.";			
RT	Proc Natl Acad Sci U S A. 97:4689-4694(2000).			
CC	-!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH			

Q9rlv6 mus musculus
Q63180 rattus norv
Q9puk1 homo sapien
Q35227 mus musculus
P34182 crotalus at
Q923w9 mus musculus
Q13444 homo sapien
Q9h2u9 homo sapien
Q28475 macaca fasc
P30403 agkistrodon
Q42596 xenopus lae
Q75078 homo sapien

CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1691-GLU-1-LEU-1692
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC MATRIX.
CC -1- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY
CC INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY
CC LIPOPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA
CC CELLS OF PREOVULATORY FOLLICLES.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 7.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB001735; BAA24501.1; ALT_INIT.
DR EMBL; D67076; BAA11088.1; ALT_FRAME.
DR MEROPS; M12.222;
DR MGD; MGI:109249; Adamts1.
DR InterPro; IPRO01762; Disintegrin.
DR InterPro; IPRO02870; Pep_M12B_propep.
DR InterPro; IPRO01590; Reprolysin.
DR InterPro; IPRO00884; TSPI.
DR InterPro; IPRO00130; Zn_MTpeptdse.
DR Pfam; PF00090; tsp_1; 3.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR SMART; SM00209; TSPI; 3.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS00092; TSPI; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR Repeat; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Hydrolase; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 48
FT PROPEP 49 253
FT CHAIN 254 968
FT SITE 206 206
FT METAL 402 402
FT ACT_SITE 403 403
FT METAL 406 406
FT METAL 412 412
FT DOMAIN 477 559
FT DOMAIN 560 617
FT DOMAIN 618 725
FT DOMAIN 726 850
FT DOMAIN 851 909
FT DOMAIN 910 968
FT DOMAIN 195 199
FT CARBOHYD 721 721
FT CARBOHYD 721 721
FT CARBOHYD 765 765
FT CARBOHYD 783 783
FT CARBOHYD 946 946
FT MUTAGEN 403 403
FT CONFLICT 335 335
FT CONFLICT 425 425
SQ SEQUENCE 968 AA; 105841 MW; 42EBDA55499FB6C1 CRC64;

Query Match 49.4%; Score 1562.5; DB 1; Length 968;
Best Local Similarity 50.4%; Pred. No. 3.2e-112;
Matches 315; Conservative 79; Mismatches 154; Indels 77; Gaps 16;
QY 1 MLLGILTLFAGRTAGG--FEPEREVVVPIRLDPDINGRRYYWRGP-EDSGDGLFIQI 57
DB 37 LLLASITMLLCARGAHRPTDEELVLP-SLE-----RAPGHDSITTRL--RL 83
QY 58 TAFQDEFLHTLTPDAQFLAPAFSTEHLGVPLQGLTGG--SDLRRCFFSGDVNA 109
DB 84 DAFGQLLKLQPDGGLAPGFTLTQV-----GRSPGEAQHLDPDGLAHCFSGVNG 138
QY 110 EPDFAAYSLCGLRGAFYRGAEYVVISPLPNAS-----APAAQRNSQGA-----HLLQRRG 161
DB 139 DPGSAALSLCEGVRGAFYLGEEFFIQAPGVATERLAPAVPEESSARPOFHILRRR- 197
QY 162 VPGSGDPTSRGCVASGWNPAIRALDPYKPRRAGGESRRR----- 206
DB 198 ---RRSGGAKCGYMD-----DETLPSTDSRPSQNTNRNQPVRDPTPDAGKP 243
QY 207 ---SGRAKRFVSIPIRYVETLWADSMVKPHGADLEHYLLTLAARLYRHPISILNPI 262
DB 244 SGPSIRKRFVSPRYVETLWADSMVADFGSLGKHYLLTFSVAARFYKHFSIRNSI 303
QY 263 NIVYKVLRLDRDGPVKVTGNAALTLRNFCAMQKLNKYSKHPYWDYTAILEFTRODLC 322
DB 304 SLVYKVIIVYEEKGPEVTSNAALTNRNFCNQKNSPSDRDPEHYDYTAILEFTRODLC 363
QY 323 GATTCDDTIGMADVGMCDPKSCSVIEDDGLPSAFTTAHELGHVFNPHDMNVKCEVFG 382
DB 364 GSHTCDDTIGMADVGMCDPKSCSVIEDDGLQAQAFSTAHELGHVFNPHDDAKKASLNG 423
QY 383 KLRANHMSPTLIQIDRANPWSAGSAIITDLDGSHGDCLLDQPSKPISLPDLPGASY 442
DB 424 VTGDSHLMASLLSDHSPWSPCSAYMVTSELDNGHGECLMDKRPQPIKLPSDLPCTLY 483
QY 443 TLSQOCELAFGVGSKPCP-YMQYCTKLWCTGKAGQMVQCTRHFPFWADGTSCGEGKCLCK 501
DB 484 DANROCFTFGEESKHCPDAASTCTTLWCTSGLLVCQTKHFPWADGTSCGEGKWCVS 543
QY 502 GACVERHNLNKH---RVDGSAKWADPYGPCSRTCGGVQVLARROCTNPTPANGKYGEGV 558
DB 544 GCVNKTDM-KHATPVHGSWGPWGDCSRTCGGVQVYTMRECDNFVPKNGKCYCEGK 602
QY 559 RVKYRSCNLEPCPSASGSKSFREQ 583
DB 603 RVRYSRSCNIEDCPDN-NGKTFREQ 626
RESULT 2
ATSL_HUMAN
ID ATSL_HUMAN STANDARD; PRT; 967 AA.
AC Q9UH18; Q9UP80; Q9UH83; Q9P2K0; Q9NSJ8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1) (METH-1).
OS ADAMTS1 OR METH1 OR KIAA1346.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Casas C., Pritchard M.A., Estivill X., Arbones M.L.;
FT "Cloning, characterization and mapping on human chromosome 21 of the
FT orthologue of murine Adams-1.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A., AND FUNCTION.
RP TISSUE=Heart;
RX MEDLINE=99367466; PubMed=10438512;

[illegible]

```
QY 328 DTGLMADVCTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNNKVVCEVEFGKLRAN 387
DQ 368 DTGLMADVCTMCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNNKVVCEVEFGKLRAN 427
QY 388 HWSMPTLIDRANPWSACSAAITDFDSDSHGDCDLDQPSKPTSLPEDLPAGASYTLISQ 447
DQ 428 HWSMPTLIDRANPWSACSAAITDFDSDSHGDCDLDQPSKPTSLPEDLPAGASYTLISQ 487
QY 448 CELAFGVSGKPCP-YMQVCTKLWCTKAKGQMVQTRHFWADGTSGEGKLCGLKACVE 506
DQ 488 COFTGEDSKHCPDAASTCTWCTGTSGGLVQCTHFWADGTSGEGKLCGLKACVE 547
QY 507 RNLNKKH--RVDSGAKWDVPGPCSRCTCGGGVOLARQCTNPPANGKGYCEGVRYKYR 563
DQ 548 KTD-RKHEDTPHSGWGWGFWGDCSRCTCGGGVOLARQCTNPPANGKGYCEGVRYKYR 606
QY 564 SCNLPCPSSASGSKFREQ 583
DQ 607 SCNLPCPSSASGSKFREQ 625

RESULT 3
ATSL_RAT
ID ATSL_RAT STANDARD; PRT; 967 AA.
AC Q9W0Q1; Q9ER11;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-1 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 1) (ADAM-TS 1) (ADAM-TS1).
GN ADAMTS1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Brain;
RA Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,
RA Little S.P.;
RA "Induction of a disintegrin and metalloproteinase with the
RA thrombospondin type I motif (ADAMTS).";
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 18-967 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Liver;
RX MEDLINE=20304099; PubMed=10847486;
RA Diamantis I., Luetli M., Hoesli M., Reichen J.;
RT "Cloning of the rat ADAMTS-1 gene and its down regulation in
RT endothelial cells in cirrhotic rats.";
RL Liver 20:165-172(2000).
CC -!- FUNCTION: CLEAVES AGGREGAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC SIMILARITY). ACTIVE METALLOPROTEINASE, WHICH MAY BE ASSOCIATED WITH
CC VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: CLEAVES AGGREGAN AT THE 1683-GLU(-)-LEU-1684
CC SITE, WITHIN THE CHONDROITIN SULFATE ATTACHMENT DOMAIN.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM
CC CIRRHOTIC LIVER.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FUIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF149118; AAC34012.1; -
CC EMBL; AF304446; AAG29823.1; -
CC MEROPS; M12.222; -
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR002870; pep_M12B_propep.
CC InterPro; IPR001590; Repolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; za_Mtpeptdse.
CC Pfam; PF00090; tsp_1; 6.
CC Pfam; PF01421; Repolysin; 2.
CC Pfam; PF01562; Pep_M12B_propep; 2.
CC SMART; SM00209; TSP1; 3.
CC PROSITE; PS00142; ZINC_MEPPO; 1.
CC PROSITE; PS0092; TSP1; 2.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC Repeat; Extracellular matrix; Zinc; Signal; Glycoprotein; Zymogen;
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Zinc; Signal; Glycoprotein; Zymogen;
FT SIGNAL 1 54
FT PROPEP 55 252
FT CHAIN 253 967
FT SITE 205 205
FT METAL 401 401
FT AC_SITE 402 402
FT METAL 405 405
FT METAL 411 411
FT DOMAIN 476 558
FT DOMAIN 559 615
FT DOMAIN 616 724
FT DOMAIN 725 857
FT DOMAIN 858 907
FT DOMAIN 908 967
FT DOMAIN 194 198
FT CARBOHYD 547 547
FT CARBOHYD 720 720
FT CARBOHYD 764 764
FT CARBOHYD 782 782
FT CARBOHYD 945 945
FT CONFLICT 21 21
FT CONFLICT 26 31
FT CONFLICT 49 49
FT CONFLICT 72 72
FT CONFLICT 79 79
FT CONFLICT 249 249
FT CONFLICT 262 265
FT CONFLICT 607 607
FT CONFLICT 936 936
FT CONFLICT 962 962
SQ SEQUENCE 967 AA; 105705 MW; F93C864F6DCB4CF CRC64;

Query Match 48.9%; Score 1547; DB 1: Length 967;
Best Local Similarity 50.9%; Pred No 5e-111;
Matches 312; Conservative 82; Mismatches 165; Indels 54; Gaps 15;

QY 1 MLLGLITLAFAGRTAGG--FEPEREVVVPIRLDPDINGRRYYWRGPDGSGDGLIFQIT 58
DQ 37 LLLLSITMLLCVGRGHPTEDEELVL-----PSLERARGH-----DSTTLRLD 83
QY 59 AFQEDFYHLHLPDAQFLAPAFSTHGLVPLGLTGGG-----SDLRRCFVSGDVNAE 110
DQ 84 AFGQLHLKLPDGSGLAPAGFTLV-----GRSPGSEAOHLDPDGLAHCFFYSGVNGD 138
QY 111 PDSFAAVSLCGLRGATCYGAEYVISPPLNPNAS-----APAAQRNSQGA-----HLL--QRR 160
DQ 139 PSSAAALSCGVGAFYVLCQEEFFIQAPAVATERLVPAEPKEESIAPPRFHILRRRR 198
```

```
QY 161 GYPGPGSD-----PTSRGVASGWNPAIRLALDPYKPRRAGFGESKRRSRGRKRFV 214
Db 199 GSGGAKCGVDETLPTNSGRESQNTPDQWPLNP-TPQAG----KPTGPGSRKRRFV 254
QY 215 SIPRVETLVVADSWKFGHADLEHYLLTLLATAARLYRHPSILNPNINIVVKVLLLRD 274
Db 255 SSPRVETMLVADQSMADFHGSLGKHYLLTFLSVAARFYKHPISIRNSISLVVKKILYIE 314
QY 275 RDSGPKVTGNALTLRNFCAWKLNKVSQKHPEYWDATILFTQDLGCGATTCDTLGMAD 334
Db 315 EQKPEVTSNAALTLRNFCSQKHNSPDRDPEDHYDTAILFTQDLGCGSHSTCDTLGMAD 374
QY 335 VGTMDPKRSCSVIEDDGLPSAFTTAHELGHVFNHNDNVKCEBEVFGKLRANHHMSPIL 394
Db 375 VGTCDPSRSCSVIEDDGLQAAFTTAHELGHVFNHNDNVKCEBEVFGKLRANHHMSPIL 434
QY 395 IQIDRANPWSACSAALITDFLDGSHGCLLDQSPKPSILPEDLPAGSYTILSQQCEALFV 454
Db 435 SSLDHSQWPCSAAYMYTSLFNDHGHECLMDKPNQIKLPDLPGLTLYDANROCOQFTFE 494
QY 455 GSKPCP-YMQYCTKLTCTGKAKGOMVQOTRHFHPWADGTCGEGKCLKLGACVERHNLKH 513
Db 495 ESTHCPDAASTCTLTWCTGTSGGLVLCQTKHFPWADGTCGEGKWCVCVKNTDM-KH 553
QY 514 ---RVDSWAKWDYPGCSRTCCGGVQLARRQCTNPTPANGKYCEGVRRKYRSCNLEPC 570
Db 554 FATPVHGSWGPWGPWDCSRTCCGGVQYTMRECDNPVPKNGKYCEGKRYRSCNIEDC 613
QY 571 PSSAGKGSFREQ 583
Db 614 PDN-NGKTFREQ 625

RESULT 4
ATS8_HUMAN STANDARD; PRT; 890 AA.
AC Q9UP79; Q9NZS0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ADAMTS-8 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 8) (ADAM-TS 8) (ADAM-TS8) (METH-2)
DE (METH-8)
GN ADAMTS8 OR METH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=99367466; PubMed=10438512;
RA Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S., Lombardo M., Iruela-Arispe M.L.;
RT "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new family of proteins with angio-inhibitory activity.";
RL J. Biol. Chem. 274:23349-23357(1999).
RN 2;
RP SEQUENCE OF 195-440 FROM N.A.
RX MEDLINE=20079168; PubMed=10610729;
RA Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on mouse chromosome 9 and human chromosome 11.";
RL Genomics 62:312-315(1999).
CC -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND KIDNEY.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
```

```
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FUFIN ENDOPEPTIDASE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF060153; AAD48081.1; -.
CC EMBL: AF175283; AAF25806.1; -.
CC HSSP: P34179; 1IAG.
CC MEROPS: M12.226; -.
CC Genew: HGNC:224; ADAMTS8.
CC MIM: 605175; -.
CC InterPro: IPR001762; Disintegrin.
CC InterPro: IPR002870; Pep_M12B_propep.
CC InterPro: IPR001590; Reprolysin.
CC InterPro: IPR000884; TSPL.
CC InterPro: IPR000130; Zn_MTpeptdse.
CC Pfam: PF00090; tsp_1; 2.
CC Pfam: PF01421; Reprolysin; 1.
CC Pfam: PF01562; Pep_M12B_propep; 1.
CC SMART: SM00209; TSPL; 2.
CC DR PROSITE: PS00142; ADAM_MEPRO; 1.
CC DR PROSITE: PS00142; ZINC_PROTEASE; 1.
CC DR PROSITE: PS00092; TSPL; 1.
CC DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 27
FT PROPEP 28 214
FT CHAIN 215 890
FT METAL 364 364
FT AC1_SITE 365 365
FT METAL 368 368
FT METAL 374 374
FT DOMAIN 439 526
FT DOMAIN 527 583
FT DOMAIN 584 690
FT DOMAIN 691 832
FT DOMAIN 833 890
FT DOMAIN 202 205
FT CARBOHYD 345 345
FT CARBOHYD 401 401
FT CARBOHYD 466 466
FT CARBOHYD 491 491
FT CARBOHYD 600 600
FT CONFLICT 195 195
FT CONFLICT 413 440
SQ SEQUENCE 890 AA; 96671 MW; 57D70E03D5739D3 CRC64;
Query Match 43.18; Score 1362; DB 1; Length 890;
Best Local Similarity 45.4%; Pred. No. 7.2e-97;
Matches 284; Conservative 81; Mismatches 171; Indels 90; Gaps 18;
QY 1 MLLGILALFA----GRTAGGFEPEVVPVPIRLDPDINGRRYWRGPDGGLIFQ 56
Db 14 LLLLLLLLLPLARGAPARPAAGGAS-ELVVPTEL-----PGSAGE--LALH 57
QY 57 ITAQEDYHLHTDPAQAFAPAFSTEHLGVLPLQGLTGSSDLRRCFYSGDVNAEPDSFAA 116
Db 58 LSATGKGVLRAPDSDSLAPEFKIERLGSGRA-TGGERGLRCFFSGTNGEPESIAA 116
QY 117 VSLGGLRGAGCYRCAEVVISPLNAPASPAORNSQGAHLRQGVGGSGDPTSRGCV 176
Db 117 VSLKGLSGSFLDGEETIOP-QGAGGSLAQ-----PHRLQR----- 153
```

```
QY 177 ASGNPAILRALDPYKPR---RAGFESRSR-----RRS 207
DB 154 ---WGAPAGAPL-PRGPEVETGEGORGDHQEDSEESQEEAEAGASEPPPLGAT 209
QY 208 GRAKREVSIPRYVETLIVVADESVMKPHGADLEHYLLTLTAARLYRHPISILNPINIVV 267
DB 210 SRTKREVSERFVETLLVADASMAAPYGADLQNHILTLMSVAARYYKHPISIKNSILMWV 269
QY 268 KVLILLDRDGSQVATGNAALTNRNFCWOKKLNKYSDKHPEYWDTAIFLTRODLGAT-T 326
DB 270 KVLIVDEKWKPEVSDNGGLTNFCWQRNFQPSDRPEHYDTAILLTRNFCGQEGCL 329
QY 327 CDTLGMADVTCMDPKRRSCSVIEDGLPSAFTTAHELGHVFNHNDVKNVCEVFGLRA 386
DB 330 CDTLGVADIGTICDPNKSVCVIEDEGLQAHTLAHELGHVLSMPHDDSKPCTRLFGPMGK 389
QY 387 NHMSPFTLQIDRANPWSACSAIITDFLDGSHGDCLLDQPKPISLPEDLP--ASYTL 444
DB 390 HHVNAFLFVHLNQTLPWSPCSAMYLTELLDGGHGDCLLDAPGAALPLPTGLPGRMALYOL 449
QY 445 SQCELAFGVGSKPCP---YMQVCTKLWCTGKRAKGQMVQTRH--PPWADGTSCGEGKLC 499
DB 450 DQCRQIFGPDFHCNPTNSQDVCAQLWCHTDG-AEPLCHTKNGSLPWADGTPCGPGHLC 508
QY 500 LKACVVERHNLNKH--VDGSAWKADPYPCSTCGGCVOLARRQCTNPTPANGKYCEG 557
DB 509 SEGSCLPPEEVERPKPVVDGWAAPWGPWBCSRTCGGVQVFSHRECKDPEPQNGGRYCLG 568
QY 558 VRKYRSCNLEPCSSASGHSFREQ 583
DB 569 RRAKYQSCHTTECP--PDGKSFREQ 592

RESULT 5
AT54_HUMAN STANDARD; PRT; 837 AA.
ID AT54_HUMAN
AC O75173; Q9UN83;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-4 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 4) (ADAM-TS 4) (ADAM-TS4) (Aggrecanase 1)
DE (ADMP-1).
GN ADAMTS4 OR KIAA0688.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99286303; PubMed=10356395;
RA Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
RA Liu R.-Q., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
RA Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,
RA Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
RA Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
RA Trzaskos J.M., Arner E.C.;
RT "Purification and cloning of aggrecanase-1: a member of the ADAMTS
RT family of proteins.";
RL Science 284:1664-1666(1999).
RN [3]
RP SEQUENCE FROM N.A.
```

```
RA Sawaji Y., Nagase H., Saklatvala J., Clark A.R.;
RL "ADAMTS-4 genomic locus.";
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
[4]
RN PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX MEDLINE=20400518; PubMed=10827174;
RA Tortorella M.D., Pratta M.A., Liu R.-Q., Abbaszade I., Ross H.,
RA Burn T.C., Arner E.C.;
RT "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
RT aggrecan substrate recognition and cleavage.";
RL J. Biol. Chem. 275:25791-25797(2000).
CC -!- FUNCTION: CLEAVES AGGECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC DESTRUCTION OF AGGECAN IN ARTHRITIC DISEASES. COULD ALSO BE A ONE
CC CRITICAL FACTOR IN THE EXACERBATION OF NEURODEGENERATION IN
CC ALZHEIMER'S DISEASE.
CC -!- CATALYTIC ACTIVITY: Cleaves aggrecan at the 392-Glu-|-Ala-393
CC site.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (by similarity).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED
CC AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
CC -!- INDUCTION: BY INTERLEUKIN-1.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO AS ADAMTS2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB014588; BAA31663.1; -
CC EMBL; AF148213; AAD1494.1; -
CC EMBL; AY044847; AAL02262.1; -
CC MEROPS; M12.221; -
CC Genew; HGNC:220; ADAMTS4.
CC MIM; 603876; -
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR001590; Reprolysin.
CC InterPro; IPR000884; TSP1.
CC InterPro; IPR000130; Zn_Mtpeptdse.
CC Pfam; PF00090; tsp_1; 1.
CC Pfam; PF01421; Reprolysin; 1.
CC SMART; SM00209; TSP1; 1.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC PROSITE; PS0215; ADAM_MERO; 1.
CC PROSITE; PS50092; TSP1; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Extracellular matrix.
KW SIGNAL 1 51 POTENTIAL.
FT PROPEP 52 212
FT CHAIN 213 837
FT SITE 194 194
FT METAL 361 361
FT ACT_SITE 362 362
FT METAL 365 365
FT METAL 371 371
FT DOMAIN 437 519
FT DOMAIN 520 576
FT DOMAIN 577 685
FT DOMAIN 686 837
FT DOMAIN 247 252
FT CARBOHYD 68 68
ADAMTS-4.
CYSTEINE SWITCH (POTENTIAL).
ZINC (CATALYTIC) (BY SIMILARITY).
BY SIMILARITY.
ZINC (CATALYTIC) (BY SIMILARITY).
ZINC (CATALYTIC) (BY SIMILARITY).
DISINTEGRIN-LIKE.
TSP TYPE-1.
CYS-RICH.
SPACER.
POLY-ALA.
N-LINKED (GLCNAC. . .) (POTENTIAL).
```

CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT LUNG AND HEART AND LOW EXPRESSION DURING MOUSE DEVELOPMENT.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC EMBL: AF175282; AAF25805.1; -
CC HSSP: P34179; 1IAG.
CC MEROPS: M12.226; -
CC MGD: MGI:1353468; Adamts8.
CC InterPro: IPR001762; Disintegrin.
CC InterPro: IPR001590; Reprolysin.
CC InterPro: IPR000884; TSP1.
CC InterPro: IPR000130; Zn_MTPeptidse.
CC Pfam: PF00090; tsp_1; 2.
CC Pfam: PF01421; Reprolysin; 1.
CC SMART: SM00209; TSP1; 2.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC PROSITE: PS50215; ADAM_MEPRO; 1.
CC PROSITE: PS50092; TSP1; 1.
CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 28
FT PROPEP 29 228
FT CHAIN 229 905
FT METAL 378 378
FT ACT_SITE 379 379
FT METAL 382 382
FT METAL 388 388
FT DOMAIN 453 541
FT DOMAIN 542 598
FT DOMAIN 599 705
FT DOMAIN 706 847
FT DOMAIN 848 905
FT CARBOHYD 415 415
FT CARBOHYD 480 480
FT CARBOHYD 506 506
FT CARBOHYD 615 615
FT SEQUENCE 905 AA; 98879 MW; 124D4132B33A0CAE CRC64;
Query Match 42.0%; Score 1329.5; DB 1; Length 905;
Best Local Similarity 44.9%; Pred. No. 2.3e-94;
Matches 284; Conservative 85; Mismatches 175; Indels 89; Gaps 17;
QY 2 LLLGLITL-----AFAGRTAGGFEPEVVPVIRLDPDINGRRYYWRGPDGSGDGL 53
DB 13 LLLLQLPLPPPLVCGAPAGPTGA--QASELVVPTL-----PGSASE--L 55
QY 54 IFQITAFQDFVHLTPDAQFLAPAFSTEHLGVPLQGLTGGSSDLRCFYSGDVNAEPDS 113
DB 56 AFHLGAGGQGVFLRLAPDASFLAPEKFLERLG--GSSAAAGGEPGLCFSGTVNGERES 114
QY 114 FAVSLGCLGRGAFGYRGAEEYVISPAPNAQAQNSOGA-----HLLQRRGVPGGP 166
DB 115 LAAMSCVAGWSGFLLAGEEFTIQP-----QAGDSLDQPHRLQWRG--PGQR 160
QY 167 SGDPTSRGCVASGWNPAIRLALDPYKPRRAGFGEGRS-----RRRS----- 208


```
Db 397 IEDDGLHAAFTVAHEIGHLLGLSHDDSKFCENFQTTDKRLMSILLTSIDASKPWSKCT 456
QY 408 AAILTDFLSDGHCCLLPQKPSLIPEDLPASVTLISQOCCLAFGVSKPCPYWQYCTK 467
Db 457 SATITFEDDGHGNCCLLPKRLGLPELPQGYDATQCCNLTFGPSEYSCVPCMDYCAR 516
QY 468 LWCTCKAKGQVQCOTRHFPMADGTCGSGKGLCLAGACVER-----HNLNKHVRVDGSAKW 522
Db 517 LWCAVVRQGVQVCLTKKLPVAVGEGPCGKGRVCLQKGVKTKKKYKYSTSSH---GNNGSW 573
QY 523 DYPGCSRTCCGGVQLARRQCTNTPANGGKYCGVRVYKSCMLEPCPSASGKSPREE 582
Db 574 GPMGQCSRSCGGVGFAYRHCNCPAPRNSGRYCTGKRAIYRSCSVTPCP--PNGKSPRHE 631
QY 583 Q 583
Db 632 Q 632

RESULT 8
AT99_HUMAN STANDARD; PRT; 1629 AA.
AC Q9P2N4; Q9NR29;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE ADAMTS-9 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 9) (ADAM-TS 9) (ADAM-TS9).
GN ADAMTS9 OR KIAA1312.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE=Fetal;
RX MEDLINE=20396138; PubMed=10936055;
RA Clark M.E., Kellner G.S., Turbeville L.A., Boyer A., Arden K.A.,
RA Maki R.A.;
RT "ADAMTS 9, a novel member of the ADAM-TS/Metallospodin gene
RT family.";
RL Genomics 67:343-350(2000).
RN [2]
RS SEQUENCE OF 159-1629 FROM N.A. (LONG ISOFORM).
RC TISSUE=Brain;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirotsawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN ALL FETAL TISSUES.
CC EXPRESSED SLIGHTLY IN ADULT OVARY, PANCREAS, HEART, KIDNEY, LUNG,
CC PLACENTA. ALSO DETECTED IN SPINAL CORD AND BRAIN. NOT DETECTED IN
CC COLON, SMALL INTESTINE, TESTIS, LIVER, SKELETAL MUSCLE, SPLEEN OR
CC THYMUS.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 11 TSP TYPE-1 DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF261918; AAF89106.1; -
CC EMBL: AB037733; BAA92550.1; -
CC HSP: P15167; IATL.
CC MEROPS: M12.021; -.
CC Genew: HGNC:13202; ADAMTS9.
CC MIM: 605421; -.
CC InterPro: IPR001762; Disintegrin.
CC InterPro: IPR002870; Pep_M12B_propep.
CC InterPro: IPR001590; Reprolysin.
CC InterPro: IPR000884; TSP1.
CC InterPro: IPR000130; Zn_M12peptdse.
CC Pfam: PF01421; Reprolysin; 1.
CC Pfam: PF01562; Pep_M12B_propep; 1.
CC SMART: SM00209; TSP1; 12.
CC PROSITE: PS00215; ADAM_MEPRO; 1.
CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
CC PROSITE: PS00092; TSP1; 9.
CC PROSITE: PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix; Alternative splicing.
FT SIGNAL 1 18
FT PROPEP 19 287
FT CHAIN 288 1629
FT DOMAIN 509 587
FT DOMAIN 589 642
FT DOMAIN 645 752
FT DOMAIN 753 880
FT DOMAIN 999 1053
FT DOMAIN 1056 1108
FT DOMAIN 1111 1156
FT DOMAIN 1184 1239
FT DOMAIN 1240 1295
FT DOMAIN 1332 1383
FT DOMAIN 1386 1439
FT DOMAIN 1445 1498
FT DOMAIN 1501 1554
FT DOMAIN 1562 1612
FT DOMAIN 88 96
FT SITE 223 223
FT METAL 434 434
FT ACT_SITE 435 435
FT METAL 438 438
FT METAL 444 444
FT CARBOHYD 112 112
FT CARBOHYD 135 135
FT CARBOHYD 271 271
FT CARBOHYD 749 749
FT CARBOHYD 840 840
FT CARBOHYD 1213 1213
FT CARBOHYD 1267 1267
FT VARSPIC 1064 1072
FT VARSPIC 1073 1073
FT CONFLICT 367 367
SQ SEQUENCE 1629 AA; 182649 MW; C1C4CEFF58B941F CRC64;
Query Match 39.0%; Score 1232.5; DB 1; Length 1629;
Best Local Similarity 40.6%; Pred. No. 1.3e-86;
Matches 252; Conservative 87; Mismatches 210; Indels 71; Gaps 13;
QY 22 EREVVVPIRLD-----DINGRRYY-----RGPEDSGDGLIFQITAFQ 61
Db 47 EYEIVSPIRVNALGPEPTNVHFKRTRESINATDPWAFASSSSSTSSQAHRLSAFG 106
QY 62 EDFYHLPTDQAFAPASTHEHLGVP-----LQLTGGSSDLRRFCYGDVNAEPDPAFV 117
Db 107 QQFLNLTANAGFIAPLFTVILGLTGPVGNQTKFYSEEAELKHCYKGYVNTNHTAVI 166
```


entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC Query Match 36.9%; Score 1168.5; DB 1; Length 630;
CC Best Local Similarity 54.9%; Pred. No. 3.4e-82;
CC Matches 208; Conservative 52; Mismatches 114; Indels 5; Gaps 4;
CC
DR 209 RAKRFVSIPIRYVETLVVADSMVKFAGADLEHYLLTLLATAARLYRHPHSILNPINIVVYK 268
DR EMBL; AF140675; AAD56358.1;
DR HSP; P15167; IATL.
DR MEROPS; M12.231;
DR Genew; HGNC:223; ADAMTS7.
DR MIN; 605009;
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF00090; tsp_1; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR SMART; SM00209; TSP1; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS02115; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSP1; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 27
FT PROPEP 28 232
FT CHAIN 233 997
FT SITE 204 204
FT METAL 388 388
FT ACT_SITE 389 389
FT METAL 392 392
FT METAL 398 398
FT DOMAIN 462 537
FT DOMAIN 538 594
FT DOMAIN 595 697
FT DOMAIN 698 914
FT DOMAIN 915 990
FT CARBOHYD 94 94
FT CARBOHYD 693 693
FT CARBOHYD 778 778
SQ SEQUENCE 997 AA; 109694 MW; 6587044ED02FC104 CRC64;
Query Match 29.6%; Score 937.5; DB 1; Length 997;
Best Local Similarity 39.4%; Pred. No. 3.2e-64;
Matches 250; Conservative 69; Mismatches 216; Indels 99; Gaps 29;
QY 1 MLLILGILTLAFAGRTAGGFEPEP---EVVPIRLDP-----DINGRR- 39
DB 18 LLLICALAPGAPGAPGRATEGRAALDIVHPVRVDAGGSFSLYELWPRALRKRDVSVRRD 77
QY 40 ---YY---WRPEDSDQGLIQTAFQEDFYHLTPDAQFLAPAF---STEHLGVPLQGL 91
DB 78 APAFYELQYRGRE-----LRFNLTAQ-----HL-----LAPGVSETRRRG---GL 116
QY 92 TGGSSDLR---RCFYSGDVNAEPD---SFAAVSLCGGLRGAGFYRGAEYVISPINASA 144
DB 117 --GRAHRAHTFACHLLGEVQ-DPELEGLAIAISACDGLKGVFQLSNEDYFIEPLD--SA 171
QY 145 PAAQRNSQGAHLQRRGVP-----GGPSGDPTSRGCVASGWNPAILLALDPKPKPRAGFG 199
DB 172 PARPGHAQ-PHVYKRAQPERLAQRGDSSAP-STCGV-----QVYELESRRERW 219
QY 200 ESRSRRSRGRKRF-----VSIPRYVETLVVADESMVKPHG-ADLEHYLLTLLATAARLYR 254
DB 220 EQRQWRRLRLRLHORSVSKKVVETLVVADAKMVEYHGQVQSVSYVLTINMVAAGLFH 279
QY 255 HPSILNPINIVVVKVLLLRDRDSGPKVTGNAALTIRNFCWQKLNKYSKHPYWDTAI 314
DB 280 DPSIGNPIHTIVRVLLLEDEEDLKITHADNTLSKFCWKQKSNMKGDAHPLHDTAI 339
QY 315 LFTRODLCCGATT--CDTLGMADVGMCMCDPKRCSVIEDDGLPSAFTTAHELGHVFNMPHD 372
DB 340 LLTRKDLCAAMNRPCTETGLSHVAGMCQPHRSCINEDTGLPLAFTVAHELGHSEGIQHD 399
```

```
QY 209 RAKRFVSIPIRYVETLVVADSMVKFAGADLEHYLLTLLATAARLYRHPHSILNPINIVVYK 268
DB 2 RTRKRFASLSRFVETLVVADDKMAAFHGAGLKHYLLTVMAAAAKAFKPSIRNPVNLVTR 61
QY 269 VLLLRDRDSGPKVTGNAALTIRNFCWQKLNKYSKHPYWDTAI LFTRODLCCGATTCD 328
DB 62 LVILGSGQEVPGVPGSAQAQTLRSFCTWQKLNPNPNDSDPHEDTAI LFTRODLCCGVTSCD 121
QY 329 TLGMADVGMCMCDPKRCSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEEVFGK-LRAN 387
DB 122 ALGMAGVTVCDPARSCAIVEDDGLQSAFTAHELGHVFNMLHDSKPCANLNGQSSSR 181
QY 388 HMSPTLIQIDRANPSACSAAIITDLSGGHDCILLDPQSPKIPISIPEDLPASVYTLSSQ 447
DB 182 HVMAVPMAHVDPEEPWSPCAREITDLDNGYGHCLLDKPEAPLHLPVTFPGKDYDADRQ 241
QY 448 CELAFGVGSKPCVM-QYCTKLWCTGKAKGMVQCTRHPPWADGTSCGGKCLKAGACVE 506
DB 242 COLTFGPDSSHCQLPPLPPCAALMCFGLNHGAMCQTKHSPWADGTCGGAQAQMGRCILH 301
QY 507 RHNLNKHVRD--GSWAKWDYGPCSRCTCGGGVQLARROCTNPTPANGKYCEGVRYKYS 564
DB 302 VDQLKDFNIPQAGGWGPMGPGDCSRCTCGGGVQFSSRDCTKPVPRNGKYCEGRRTPFRS 361
QY 565 CNLEPCPSSASGKSFREEQ 583
DB 362 CNTKNCP-HGSALTFFREEQ 379
RESULT 11
AT57_HUMAN STANDARD; .PRT; 997 AA.
AC Q3UKP4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMTS-7 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 7) (ADAM-TS 7) (ADAM-TS7).
GN ADAMTS7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=9395124; PubMed=10464288;
Hurskainen T.L., Hirohata S., Seidlin M.F., Apte S.S.;
"ADAM-TS5, ADAM-TS6, and ADAM-TS7, Novel Members of a New Family of
Zinc Metalloproteases";
J. Biol. Chem. 274:25555-25563(1999).
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, PLACENTA, LUNG,
LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS.
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
```

QY 373 -NVKVEEYFGKLRHNMSPPTLIQIDRANPWSACSAIITDLDHSGDCLLDOPSKP- 430
Db 400 GSGNDECPV-GR--RPFIMSPQLLYDAAPLTWSRCSROYITRFLDRHGGLCLDDPPAKDI 456
QY 431 ISLPEDLFGASTYLSQOCELAFGVGSKPCPYM-QYCTKLWCTGKAKGOMVOTRHFPPWAD 489
Db 457 IDFPSPVPGVLYDVSHOCLROYGAYSFCEDMDNVCHTLWCSVGT---TCHSKLDAAVD 512
QY 490 GTSCGEGKLCCKGACVERHNLKHKRVDSWAKWDYGPCSRCTCGGVOLARRCQTNPTPA 549
Db 513 GTRCENKWLSCGCV-PGRPEAVDGGWGSAWSICSRSGMGVOSAEQCTQTPPK 571
QY 550 NGGKYCEGVYKRSCLNEPCPSSASGKSFREQ 583
Db 572 YGRYCVGERKFRCLNLQACP--AGRPSFRHVQ 603

RESULT 12

AT10_HUMAN
AC Q9H324; STANDARD; PRT; 1077 AA.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ADAMTS-10 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
GN ADAMTS10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Apte S.S.;
RT "ADAM-TS10: a novel member of the ADAM-TS family containing multiple
thrombospondin type I repeats.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
matrix (By similarity).
CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch).

EMBL: AF163762; AAC35563.1; -
DR MEROPS; M12.235; -
DR Genew; M12.235; -
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep.M12B.propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR000130; Zn_M12peptdse.
DR Pfam; PF00090; tsp_1; 5.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep.M12B.propep; 1.
DR SMART; SM00209; TSP1; 5.
DR PROSITE; PS50215; ADAM_MEROPS; 1.
DR PROSITE; PS00092; TSP1; 2.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Glycoprotein; zymogen;
KW Repeat; Extracellular matrix.

FT NON_TER 1 1
FT PROPEP <1 207
FT CHAIN 208 1077
FT METAL 366
FT ACT_SITE 367 367
FT METAL 370 370
FT METAL 376 376
FT DOMAIN 434 520
FT DOMAIN 578 679
FT DOMAIN 680 802
FT DOMAIN 521 577
FT DOMAIN 799 860
FT DOMAIN 862 918
FT DOMAIN 922 976
FT DOMAIN 981 1031
FT CARBOHYD 64 64
FT CARBOHYD 196 196
FT CARBOHYD 297 297
FT CARBOHYD 714 714
FT CARBOHYD 769 769
FT CARBOHYD 866 866
SQ SEQUENCE 1077 AA; 118072 MW; 3914DE18DCBFF587 CRC64;
Query Match 28.2%; Score 891.5; DB 1; Length 1077;
Best Local Similarity 36.8%; Pred.No.1.2e-60;
Matches 219; Conservative 80; Mismatches 239; Indels 57; Gaps 20;

QY 24 EVVVPRLDPIGR-----RYWRGPEDSGDGLIFQTAFQEDFYHLHTPDAQF 74
Db 14 EIAFPTVDH--NGALLAFSPPPRRQRTGATASRLFYKVASPTHTFLNLTSSRL 71
QY 75 LAPAFSTEHLGVPLQGLTGSSDLRCFYSGDVNAE-PDSFAVSLCCGLRGAFVGRGAE 133
Db 72 LAGHVSEYV--TREGLAWQRAARPHCLYAGHLQGOASSHVAISTCGGLHLIVADEEE 129
QY 134 YVISPPLNASAPAAQNSOHAHLORRGVPGSPGPTSRGCV--ASGW--NPAILRALD 189
Db 130 YLIEPLHGGPKGSRSEESGPHVYKRSRLRPHLD--TACGVDEKPKWGRPWLRULK 187
QY 190 PYKPRAGGESRRRSRGRKRFVPIRYVETLVVVADESVMYKFGH-ADLEHYLLTLAT 248
Db 188 P--PPARPLGNETERGQPG-LKRSVSRERYETLVVADKMVMVAYHGRROVEQYVLAIMI 244
QY 249 AARLYRHPSTLNPINIVVVKVLLLRDSDGPKVTGNAALTNRNFCAWOKKL-----NK 301
Db 245 VAKLFQDSLSGTVNLVTRLLLTEDQPTLEITHAGKSLDSFCFKWQKSVNHHGHGNA 304
QY 302 VSDKHPEYWDTAILFTRODLG--GATTCDTLGMADVGTCDPKRSCSVIEDDGLPSAFTT 359
Db 305 IPENGVANHDTAVLITRYDICYKKNKPCGTGLAPVGGCERERSCSVNEDIGLPQAFI 364
QY 360 AHELGHVFNPHDNVVKCEVFG-----KLRAHNMSPPTLIQIDRANP--WSACSAIITD 413
Db 365 AHEIGHTFGMHDGVGNSCGARGQDPKAKLMAAHITMKT-----NPFVSSCNRDIYTS 417
QY 414 FLDHSGDCLLDQPSK-PISLPEDLPGASYTLTSSQCELAFAFGVSKPCPYMOTCTKLWCTG 472
Db 418 FLDHSGGLCLNRRPPQDFVYTVAPGQAYDADEQCRRHQHGVKSRQCKYGEVCSLWCL 477
QY 473 KAKGMVOCOTRHPFWADGTSCGEGKL-----CLKGACVERHNLKHKRVDSWAKWDYGPC 528
Db 478 KSNR---CITNSIPAAEGTLCQHTIDKWCYKRCVCP-FGSRPEGVGDGAWGPMTPWGDC 533
QY 529 SRTCGGVOLARRCQTNPTPANGKYCEGVYKRSCLNEPCPSSASGKSFREQ 583
Db 534 SRTCGGVSSSRHSDSPRTTGGKYCLGERRRHRNRHSDDCPPGS--ODFREVQ 586

RESULT 13

AT12_HUMAN
ID AT12_HUMAN STANDARD; PRT; 1593 AA.
AC P58397;
DT 15-JUN-2002 (Rel. 41, Created)

15-JUN-2002 (Rel. 41, Last sequence update)
 15-JUN-2002 (Rel. 41, Last annotation update)
 ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase with thrombospondin motifs 12) (ADAM-TS12) (ADAM-TS12).
 GN ADAMTS12.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal lung;
 RX MEDLINE=21264577; PubMed=11279086;
 RA Cal S., Arqueles J.M., Fernandez P.L., Lopez-Otin C.;
 RT "Identification, characterization, and intracellular processing of ADAM-TS12, a novel human disintegrin with a complex structural organization involving multiple thrombospondin-1 repeats.";
 RL J. Biol. Chem. 276:17932-17940(2001).
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely expressed in gastric carcinomas and in cancer cells of diverse origin.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY SIMILARITY).
 CC -1- PTM: THE PRECURSOR IS CLEAVED BY A URIN ENDOPEPTIDASE.
 CC -1- TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1 DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ250725; CAC20419.1; .
 CC Genbank: HGNC:14605; ADAMTS12.
 CC MIM: 606184; .
 CC InterPro: IPR001762; Disintegrin.
 CC InterPro: IPR002870; Pep_M12B_propep.
 CC InterPro: IPR001590; Reprolysin.
 CC InterPro: IPR000884; TSP1.
 CC InterPro: IPR000130; Zn_MTpeptidse.
 CC Pfam: PF00090; TSP_1; 6.
 CC Pfam: PF01421; Reprolysin; 1.
 CC Pfam: PF01562; Pep_M12B_propep; 1.
 CC SMART: SM00209; TSP1; 8.
 CC PROSITE: PS00215; ADAM_MEPRO; 1.
 CC PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
 CC PROSITE: PS50092; TSP1; 2.
 CC PROSITE: PS00142; ZINC_PROTEASE; 1.
 CC Hydrolase: Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 CC Repeat; Extracellular matrix.
 CC SIGNAL 1 25 POTENTIAL.
 CC PROPEP 26 240 BY SIMILARITY.
 CC CHAIN 241 1593 ADAMTS-12.
 CC DOMAIN 465 544 DISINTEGRIN-LIKE.
 CC DOMAIN 545 596 TSP TYPE-1 1.
 CC DOMAIN 597 700 CYS-RICH.
 CC DOMAIN 701 826 SPACER 1.
 CC DOMAIN 827 881 TSP TYPE-1 2.
 CC DOMAIN 886 943 TSP TYPE-1 3.

FT	DOMAIN	947	995	TSP TYPE-1 4.
FT	DOMAIN	996	1315	SPACER 2.
FT	DOMAIN	1316	1364	TSP TYPE-1 5.
FT	DOMAIN	1367	1423	TSP TYPE-1 6.
FT	DOMAIN	1426	1471	TSP TYPE-1 7.
FT	DOMAIN	1426	1471	TSP TYPE-1 8.
FT	DOMAIN	302	305	POLY-GLU.
FT	SITE	208	208	CYS-GLU. SWITCH (POTENTIAL).
FT	METAL	392	392	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT SITE	393	393	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	396	396	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	402	402	ZINC (CATALYTIC) (BY SIMILARITY).
FT	CARBOHYD	105	105	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	125	125	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	215	215	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	485	485	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	685	685	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	790	790	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	951	951	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1104	1104	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1275	1275	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1300	1300	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1320	1320	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1371	1371	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1378	1378	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1503	1503	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SEQUENCE	1593	1593	AA: 177545 MW: 079F48E63BD83A3 CRC64;
QY	21 PEREYVVPRI	106	106	Query Match 27.7%; Score 876.5; DB 1; Length 1593;
Db	50 PEYHVGVPRVDA	106	106	Best Local Similarity 36.6%; Pred. No. 2.7e-59;
QY	69 TPAQFLAPAFST	126	126	Matches 214; Conservative 92; Mismatches 230; Indels 49; Gaps 18;
Db	107 TVNQFLSNYSI	166	166	
QY	127 FGYRGAEYIS	186	186	
Db	167 FQLPHGDFP	220	220	
QY	187 ALDPTKPRAG	245	245	
Db	221 ELWREKWERH	274	274	
QY	246 LATAARLYR	305	305	
Db	275 MNVTGLFNP	334	334	
QY	306 HPEYNDTAL	363	363	
Db	335 NPVHDDAVL	394	394	
QY	364 GHVFNPHDN	422	422	
Db	395 GHSFGIQHD	451	451	
QY	423 LLDOPSKP	480	480	
Db	452 LDDIPKKG	507	507	
QY	481 QTRHFPWAD	538	538	
Db	508 RSKLDAADG	564	564	
QY	539 ARQCTNPT	583	583	
Db	565 AERLCNNP	607	607	

RESULT 14

AT2S2_HUMAN STANDARD; PRT; 1211 AA.

AC O95450.

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE ADAMTS-2 precursor (EC 3.4.24.14) (A disintegrin and metalloproteinase with thrombospondin motifs 2) (ADAM-TS2)

DE (Procollagen I/II amino-propeptide processing enzyme) (PCNPI)

DE (N-proteinase) (PC I-NP) (Procollagen N-endopeptidase)

DE (Procollagen I/II amino-propeptide processing enzyme).

GN ADAMTS2 OR PCNPI OR PCNPI.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI), AND DISEASE.

RC TISSUE-Skin;

RX MEDLINE=99347935; PubMed=10417273;

RA Colige A., Sieron A.L., Li S.-W., Schwarze U., Petty E., Wentelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W., Byers P.H., Lapiere C.M., Prockop D.J., Nussgens B.V.; "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis are caused by mutations in the procollagen I N-proteinase gene."; Am. J. Hum. Genet. 65:308-317(1999).

RT FUNCTION: Cleaves the propeptides of type I and II collagen prior to fibril assembly. Does not act on type III collagen. May also play a role in development that is independent of its role in collagen biosynthesis.

CC -! CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain alpha-1(I) at Pro-1-Gln and of alpha-1(II) and alpha-2(I) chains at Ala-1-Gln.

CC -! COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).

CC -! SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO COLLAGEN TYPE XIV (BY SIMILARITY).

CC -! SUBCELLULAR LOCATION: Secreted. Associated with the extracellular matrix (By similarity).

CC -! ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI; ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-PROCOLLAGEN PEPTIDASE ACTIVITY.

CC -! TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVEL IN SKIN, BONE, TENDON AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.

CC -! DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.

CC -! PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY SIMILARITY).

CC -! DISEASE: Defects in ADAMTS2 are the cause of Ehlers-Danlos syndrome type VIIC (EDS-VIIC), a recessively inherited connective-tissue disorder characterized clinically by severe skin fragility and joint hypermobility and biochemically by the presence in skin of procollagen incompletely processed at the N-terminus.

CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.

CC -! SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.

CC -! SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.

CC -! CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AJ003125; CAA05880.1; .

CC MEROPS; M12.301; .

CC Genew; HGNC:218; ADAMTS2.

CC MIM; 604539; .

CC MIM; 225410; .

CC InterPro; IPR001762; Disintegrin.

CC InterPro; IPR002870; Pep_M12B_propep.

DR InterPro; IPR001590; Reprolysin.

DR InterPro; IPR000884; TSPI.

DR InterPro; IPR000130; Zn_MTpeptdse.

DR Pfam; PF00090; tsp_1; 4.

DR Pfam; PF01421; Reprolysin; 1.

DR Pfam; PF01562; Pep_M12B_propep; 1.

DR SMART; SM00209; TSP1; 4.

DR PROSITE; PS00215; ADAM_MEPRO; 1.

DR PROSITE; PS00092; TSP1; 1.

DR PROSITE; PS00142; ZINC_PROTEASE; FALSE NEG.

DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.

DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.

KW Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;

KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;

KW Alternative splicing; Ehlers-Danlos syndrome.

FT SIGNAL 1 29 POTENTIAL.

FT PROPEP 30 253 BY SIMILARITY.

FT CHAIN 254 1211 ADAMTS-2.

FT METAL 408 408 ZINC (CATALYTIC) (BY SIMILARITY).

FT ACT_SITE 409 409 BY SIMILARITY.

FT METAL 412 412 ZINC (CATALYTIC) (BY SIMILARITY).

FT METAL 418 418 ZINC (CATALYTIC) (BY SIMILARITY).

FT DOMAIN 480 560 ZINC (CATALYTIC) (BY SIMILARITY).

FT DOMAIN 561 617 DISINTEGRIN-LIKE.

FT DOMAIN 618 722 TSP TYPE-1 1.

FT SITE 691 693 CYS-RICH.

FT DOMAIN 723 851 CELL ATTACHMENT SITE (POTENTIAL).

FT DOMAIN 852 911 SPACER.

FT DOMAIN 912 974 TSP TYPE-1 2.

FT DOMAIN 975 1030 TSP TYPE-1 3.

FT DOMAIN 40 43 TSP TYPE-1 4.

FT DOMAIN 185 188 POLY-GLU.

FT CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 251 251 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 949 949 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 993 993 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1031 1031 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1098 1098 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1145 1145 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 1150 1150 N-LINKED (GLCNAC. .) (POTENTIAL).

FT VARSPIC 544 566 HGFCHCIWLTDPILKRGSGWA -> FRPGVAHACYP

FT VARSPIC 567 1211 TLGGQGRWIA (IN ISOFORM SPNPI).

FT SEQUENCE 1211 AA; 134722 MW; BECEE25C23CAD2D CRC64;

Query Match 24.9%; Score 788.5; DB 1; Length 1211;

Best Local Similarity 34.9%; Pred. No. 1.1e-52;

Matches 217; Conservative 79; Mismatches 246; Indels 79; Gaps 27;

QY 9 LAFAGTAG---GFEPEREYVVPVIRLDPD-----INGRR-----YYW 42

DB 39 LAAADPPGGPLGHGAERILAVPVRTDAQRLVSHVVSAAVSAGVRRARRAAPTPTSP 98

QY 43 RGPEDSGDQLIFQITAFQEDFVHLTPDAQFLAPAFSTHEHLGVPLGLOGTSSDLRRCF 102

DB 99 GGNPEPGSHLFYVNVTFGRDLHLRPNARLVPAGATMEWQGE--KGTRVEPLLGSL 156

QY 103 YSGDVN--AEPDSFAAVSLCGGLRGAFGYRGAEYVISPPLPNASAPAAQNSOG-AHLQR 159

DB 157 YVGDVAGLAEEAS-VALSNCDSLGLIRMEEEFFIEPLEKGL--AAQAEQGRVHVYR 213

QY 160 RGVPGSPSGDPTSCRCVAGSNPAIILRALDPYKPRRAGGESSRRSRRAKRFVSIPIR 219

DB 214 RPTSPPLGGPQA---LDTG---ASLSDLSLS-RALGLVEEHANSSRRRRRAADDY 266

QY 220 -VETLVVADSMVKPHCAD-LEHYLLTLATAARLYRHPHSILNPINIVVVKLLLRDRS 277

DB 267 NIEVLGVDDSVVQFHGKEHVQKYLTLTMINIYEHDESGLAHINVLVIRILYSYGS 326

QY 278 GPKV-TGNAALTIRNFCAMQKKLNKYSVKHPEYWDITAILFTQDLCGATTCDTLGMADV 336

DB 327 MSLEIGNPSQSLENVCRWAYLQKQPDTHGDEYHDAIFLTQDF-GPSGMO--GYAPVT 383

QY 337 TWCDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNV-KVCEEVFGKLRANHMMSPTLI 395

Db 384 GMCHPVRSCTLNHEDGSSAFVVAHETGHVIGMEHDGQNRCDG---EVRIGSIMAPLVQ 440
QY 396 QIDRANPWSACSAIITDFLDGSHGDCILDPQ---SKPISLPEDLPASYTILSQOCELAF 452
Db 441 AAFHREHWSRCSQELSRYLHS--YDCLLDDPFAHDWP-ALPQ-LPLGLHYSNEQCRDF 496
QY 453 GVGSKPCPYMY---CTKLWCTGKAKGMVQOTRHPFMDGTSGBGKCLCKGACV---- 505
Db 497 GLGYMCMCTAFRTDPCKQWCS--HPDNPFCKTKGKPPDLGTWCAPKCKFGHCLWLTP 555
QY 506 ---ERHNLNKRHDGWSAKWDPYPCGSRCTCGGCVQLARRQCTNPTPANGKCYCEGVVRY 562
Db 556 DILKR-----DGSWAGSPGSCSRTCTGTVKTRQCDNPHFANGRTCSGLAYDF 607
QY 563 RSNLEPCPSSASGKSFREQ 583
Db 608 QLCSDRDCPDLSA--DFREQ 626

RESULT 15
ATSD2_BOVIN
ID ATSD2_BOVIN STANDARD; PRT: 1205 AA.
AC P79331;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ADAMS-2 precursor (EC 3.4.24.14) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 2) (ADAM-19)
DE (Procollagen I/II amino-propeptide processing enzyme) (Procollagen I
DE N-proteinase) (PC I-NP) (Procollagen N-endopeptidase) (PNPI).
GN ADAMS2 OR NPI.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=9725960; PubMed=9122202;
RA Collige A., Li S.W., Sheron A.L., Nusgens B.V., Prockop D.J.,
RA Lapierre C.M.;
RA "cDNA cloning and expression of bovine procollagen I N-proteinase: a
RT new member of the superfamily of zinc-metalloproteinases with binding
RT sites for cells and other matrix components.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).
RN [2]
RP PARTIAL SEQUENCE.
RX MEDLINE=95348096; PubMed=7622483;
RA Collige A., Beschlin A., Samyn B., Goebels Y., Van Beeumen J.,
RA Nusgens B.V., Lapierre C.M.;
RA "Characterization and partial amino acid sequencing of a 107-kDa
RT procollagen I N-proteinase purified by affinity chromatography on
RT immobilized type XIV collagen.";
RL J. Biol. Chem. 270:16724-16730(1995).
CC -!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
CC TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
CC PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
CC COLLAGEN BIOSYNTHESIS.
CC -!- CATALYTIC ACTIVITY: Cleaves the N-propeptide of collagen chain
CC at Ala-1-Gln.
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
CC COLLAGEN TYPE XIV.
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: ENZYMIC ACTIVITY IS DETECTED AT HIGH LEVEL
CC IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS
CC AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS
CC WERE DISPROPORTIONATELY HIGH IN HEART, LIVER, RETINA AND MUSCLE.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE-1 DOMAINS ARE IMPORTANT

CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -!- DISEASE: DEFECTS IN ADAMS2 ARE THE CAUSE OF DERMATOSPARAXIS, A
CC RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN
CC FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN
CC INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
CC -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMS3.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X96389; CAA65253.1; -.
CC MEROPS: M12.301; -.
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pep_M12B_propep.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR000884; TSP1.
DR InterPro: IPR000130; Zn_Mtpeptdse.
DR Pfam: PF00090; Tsp1; 4.
DR Pfam: PF01421; Reprolysin; 1.
DR Pfam: PF01563; Pep_M12B_propep; 1.
DR SMART: SM00209; TSP1; 4.
DR PROSITE: PS50215; ADAM_MEPRO; 1.
DR PROSITE: PS50092; TSP1; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; FALSE NEG.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE NEG.
KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.
FT SIGNAL 1 28
FT PROPEP 29 253
FT CHAIN 254 1205
FT METAL 402 402
FT ACT_SITE 403 403
FT METAL 406 406
FT METAL 412 412
FT DOMAIN 474 554
FT DOMAIN 555 611
FT DOMAIN 612 716
FT DOMAIN 717 845
FT DOMAIN 846 905
FT DOMAIN 906 968
FT DOMAIN 969 1024
FT SITE 585 687
FT DOMAIN 31 35
FT DOMAIN 177 180
FT CARBOHYD 104 104
FT CARBOHYD 245 245
FT CARBOHYD 942 942
FT CARBOHYD 943 943
FT CARBOHYD 987 987
FT CARBOHYD 1025 1025
FT CARBOHYD 1092 1092
FT CARBOHYD 1139 1139
FT CARBOHYD 1144 1144
SQ SEQUENCE 1205 AA; 133887 MW; 7B5B232A45320371 CRC64;
Query Match 24.5%; Score 775; DB 1; Length 1205;
Best Local Similarity 32.9%; Pred. No. 1.2e-51;
Matches 213; Conservative 82; Mismatches 246; Indels 106; Gaps 26;
QY 1 MLLGLITLTAFAGRTAG-----GPEPEVVVPIRLD----- 32
Db 16 LLLLLLPADARAAAAAAPPGGQGHGAERILAVPVRTDAQGRLVSHVVSATAPAG 75


```

QY 33 -----PDINGRRYYWRGPEDSGDGLIFOITAFQEDFYHLHTPDQAFLAPAE 81
Db 76 VTRRAAPAIQGLSG-----GSEEDGGRLFYNTVTFGRDLHLRLRNARLVAPGATV 129
QY 82 EHLGVPLQGLTGSGDLRCFYSGDV-NAEPDSFAAVSLCGGLRGAFGYRGAEYVISPLP 140
Db 130 EWQGE--SGATRVPLLGTCLYVGVAGLAESSVALSCDGLAGLIRMEEEFFIEPLE 187
QY 141 NASAPAAORNSOG-AHLQORRGP-----GGPSGDPHSRCGVASGWNPAILRALDPYKP 193
Db 188 KGL--AAKAEQGRVHVYHRTTSRPPPLGGPQALDT---GISADSLDSLSRAL----- 237
QY 194 RRAFGESRRRRSGRAKRFVSIPIRY-VETLVVADESMVKFHGAD-LEHYLLTLATAAR 251
Db 238 ---GVLEERVNSSRRMRRAADDYNIIEVLGVDSDVVPQPHGTEHVOKYLLTLMNIVNE 294
QY 252 LYRHPISILNPINIVVVKVLLLRDRSGPKV-TGNAALTLRNFCAWQKKLNKVS DKHPEY 310
Db 295 IYHDESLGAHINVLVRIILISYKGSMSLIEIGNPSQSLNVCRWAYLQKQKPDTHDEYH 354
QY 311 DTAIFLTRODLCGATCTDLGMADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVFNMP 370
Db 355 DRAIFLTRODF-GPSGMQ--GYAPVTGMCHPVRSCTLNHEDGFSFAFVVAHETGHVLGME 411
QY 371 HDNV-KVCEEVEGKLRANHMMSPTLIQIDRANPWSACSAAITDFLDSGHGDCLLDQP-- 427
Db 412 HDGQGNRCGD---EVRLGSIAPLVQAAAFHRFHSRCSQOELSRYLHS--YDCLRDDEFT 466
QY 428 -SKPISLPEDLPASYSYLSQOCELAFGVGSKPCPYMQY---CTKLWCTGKAKGMVCQTR 483
Db 467 HDWP-ALPQ-LPGLHYSNEQCRFDGLGYMMCTAFRTFPCQKLWCS-HPDNEYFCKTK 523
QY 484 HPPWADGTSCGEGKCLKAGCV-----ERHNLNKHVRDGSWAKWDPYGPCSRTC GG 536
Db 524 KGPPLDGTWCAPGKCFKGHCILWTPDILKR-----DGNWGAWSFFGSCSRTC GTG 575
QY 537 QLARQCTNPTPANGKYCEGVVKYRSCNLEPCPSSASGKSFRREQ 583
Db 576 KPTRQCDNPHANGRTCSGLAYDFQLCNSQDCPDALA--DFREQ 620

```

Search completed: April 29, 2003, 17:14:04
Job time : 11.1802 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 35.5509 Seconds
(without alignments)
3378.970 Million cell updates/sec

Title: US-10-009-332-1-copy_1_583
Perfect score: 3163
Sequence: 1 MLLGLITLAFAGRTAGGFE.....SCNLEPCPSASGKSFREQ 583

Scoring table: .BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3155	99.7	950	4 Q8TE58	Q8TE58 homo sapien
2	941.5	29.8	2165	5 Q19791	Q19791 caenorhabdi
3	865.5	27.4	1054	5 Q9W493	Q9W493 drosophila
4	803.5	25.4	1095	4 Q8TE56	Q8TE56 homo sapien
5	796	25.2	269	6 Q9GL54	Q9GL54 oryctolagus
6	793	25.1	1081	4 Q8TE60	Q8TE60 homo sapien
7	790	25.0	1688	5 Q8SX80	Q8SX80 drosophila
8	762	24.1	1072	4 Q8TE57	Q8TE57 homo sapien
9	757.5	23.9	1229	5 Q9VF61	Q9VF61 drosophila
10	753.5	23.8	1207	4 Q8TE59	Q8TE59 homo sapien
11	715.5	22.6	1223	4 Q8WX58	Q8WX58 homo sapien
12	715.5	22.6	1223	4 Q8TE55	Q8TE55 homo sapien
13	714	22.6	1159	4 Q8TE58	Q8TE58 homo sapien
14	568.5	18.0	1427	4 Q96L37	Q96L37 homo sapien
15	558	17.6	192	6 Q95N24	Q95N24 equus cabal
16	518	16.4	187	6 Q95N23	Q95N23 equus cabal

17	444	14.0	1091	5 Q9WL26	Q9WL26 drosophila
18	442.5	14.0	1062	5 Q19204	Q19204 caenorhabdi
19	433	13.7	1444	5 Q17591	Q17591 caenorhabdi
20	393	12.4	872	5 Q22580	Q22580 caenorhabdi
21	357.5	11.3	790	5 Q8T458	Q8T458 drosophila
22	335.5	10.6	899	13 Q8UVF1	Q8UVF1 coturnix co
23	332.5	10.5	479	13 Q9PWJ0	Q9PWJ0 agkistrodon
24	331	10.5	922	13 Q8UVF2	Q8UVF2 coturnix co
25	326	10.3	466	13 Q9TAX7	Q9TAX7 agkistrodon
26	322	10.2	477	13 Q9SPW2	Q9SPW2 bothrops ja
27	319.5	10.1	479	13 Q9PW78	Q9PW78 agkistrodon
28	314.5	9.9	609	13 Q9W6M5	Q9W6M5 agkistrodon
29	313.5	9.9	610	13 Q93523	Q93523 bothrops ja
30	313.5	9.9	610	13 Q8OG88	Q8OG88 bothrops in
31	310	9.8	476	13 Q9YI19	Q9YI19 agkistrodon
32	310	9.8	482	13 Q9PVK9	Q9PVK9 agkistrodon
33	309.5	9.8	610	13 Q9YI20	Q9YI20 agkistrodon
34	309	9.8	617	13 Q90499	Q90499 echis pyram
35	307	9.7	411	13 Q92031	Q92031 agkistrodon
36	307	9.7	466	13 Q9TAX6	Q9TAX6 agkistrodon
37	304	9.6	407	13 Q92032	Q92032 agkistrodon
38	303	9.6	108	6 Q9GLK6	Q9GLK6 oryctolagus
39	303	9.6	483	13 Q9IAB0	Q9IAB0 agkistrodon
40	302.5	9.6	609	13 Q90282	Q90282 crotalus at
41	300	9.5	620	13 Q42138	Q42138 agkistrodon
42	299.5	9.5	400	13 Q9IAY0	Q9IAY0 agkistrodon
43	298.5	9.4	825	11 Q8R3D3	Q8R3D3 mus musculu
44	298	9.4	117	11 Q8VHK4	Q8VHK4 mus musculu
45	297	9.4	549	13 Q90500	Q90500 echis pyram

ALIGNMENTS

RESULT 1

Q8TE58	PRELIMINARY;	PRT;	950 AA.
ID	Q8TE58		
AC	Q8TE58;		
DT	01-JUN-2002 (Tremblrel. 21, Created)		
DT	01-JUN-2002 (Tremblrel. 21, Last sequence update)		
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)		
DE	Metalloprotease disintegrin 15 with thrombospondin domains.		
GN	ADAMTS15.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE-21856482; PubMed-11867212;		
RA	Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,		
RA	Lopez-Otin C.;		
RT	"Cloning, expression analysis, and structural characterization of		
RT	seven novel human ADAMTSs, a family of metalloproteinases with		
RT	disintegrin and thrombospondin-1 domains.";		
RL	Gene 283:49-62(2002).		
DR	EMBL; AJ315733; CAC86014.1; -.		
KW	Integrin; Protease.		
SQ	SEQUENCE 950 AA; 103286 MW; 5DFBEI8285CCCC3B CRC64;		
Query Match	99.7%; Score 3155; DB 4; Length 950;		
Best Local Similarity	99.8%; Pred. No. 1.3e-253;		
Matches 582; Conservative	0; Mismatches 1; Indels 0; Gaps 0;		
QY	1 MLLGLITLAFAGRTAGGFEPEPEVVPVRLDPDINGRYYWRGPDSDGQGLIFQITAF 60		
Db	1 MLLGLITLAFAGRTAGGFEPEPEVVPVRLDPDINGRYYWRGPDSDGQGLIFQITAF 60		
QY	61 QEDFYHLHLPDAQFLAPAFSTHGLVPLQGLTGGSSDLRCRCFSGDVNAEPDSFAVSLC 120		
Db	61 QEDFYHLHLPDAQFLAPAFSTHGLVPLQGLTGGSSDLRCRCFSGDVNAEPDSFAVSLC 120		
QY	121 GGLRGAFGRGAEYVISPLPNASAPAAQNSQAHLLQRRGVPGGPGDPTSRGCVASGW 180		

```
Db 121 GGLRGAFYGAAYVISPFPNAPSAPQAQNSOGAHLQRGYPGSGDPTSCRGVASSW 180
QY 181 NPAILRALDPYKPRRAGFGESESRSSRRGRKRFVSPRVYETLVVADESMVAFHGADLEH 240
Db 181 NPAILRALDPYKPRRAGFGESESRSSRRGRKRFVSPRVYETLVVADESMVAFHGADLEH 240
QY 241 YLLTLATARLYRHSILNPINVVVYLLLRDRSDGPKVTGNALTLRNFCAWQKLN 300
Db 241 YLLTLATARLYRHSILNPINVVVYLLLRDRSDGPKVTGNALTLRNFCAWQKLN 300
QY 301 KVSQKHEPYWDTAILTRQDLGCAITCDTLGMADVGTMDPKRSCSVIEDDGLPSAFTTA 360
Db 301 KVSQKHEPYWDTAILTRQDLGCAITCDTLGMADVGTMDPKRSCSVIEDDGLPSAFTTA 360
QY 361 HELGHVFNPHDNVNVKVEEYFVKLRANHMSPITLQIDRANPWSACSAAIITDFDLSHG 420
Db 361 HELGHVFNPHDNVNVKVEEYFVKLRANHMSPITLQIDRANPWSACSAAIITDFDLSHG 420
QY 421 DCLLDQSPKIPSLPDLPGASYTLSSOCELAFGVGSKPCPYMOYCTKLWCTGKAKQWVC 480
Db 421 DCLLDQSPKIPSLPDLPGASYTLSSOCELAFGVGSKPCPYMOYCTKLWCTGKAKQWVC 480
QY 481 QTRHFPWADTSCGEGKCLKLGACVVRHNLNHRVDGWSNAKWDYPGCSRTCGGGVQLAR 540
Db 481 QTRHFPWADTSCGEGKCLKLGACVVRHNLNHRVDGWSNAKWDYPGCSRTCGGGVQLAR 540
QY 541 RQCTNPTPANGKYCGVRVYKYSNLEPCPSSASGKSFREQ 583
Db 541 RQCTNPTPANGKYCGVRVYKYSNLEPCPSSASGKSFREQ 583

RESULT 2
Q19791 ID Q19791 PRELIMINARY; PRT; 2165 AA.
AC Q19791; Q27524;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE F25H8.3 protein.
GN F25H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RA Gjadasty S.;
RP SEQUENCE FROM N.A.
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Shaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thiermy-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z69361; CAA93288.1; -.
DR EMBL; Z69360; CAA93288.1; JOINED.
DR EMBL; Z69360; CAA93287.1; -.
DR EMBL; Z69361; CAA93287.1; JOINED.
DR HSSP; P15167; IDTH.
DR MEROPS; M12.135; -.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPL.
```

```
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 14.
DR SMART; SM00209; TSPL; 18.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00092; TSPL; 6.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 2165 AA; 244397 MW; FCC3DA8AAA9C4888 CRC64;

Query Match 29.88; Score 941.5; DB 5; Length 2165;
Best Local Similarity 38.78; Pred. No. 5.3e-69;
Matches 203; Conservative 75; Mismatches 181; Indels 65; Gaps 14;

QY 101 CFYSGDVNAEPDSPAFAVSLC---GGLRGAFYGAAYVISPFPNAPS---PAAQNSQGA 154
Db 169 CIYRAHVKG-VHQHSIVNLCDSGDLGMLALPSGIHTVEPIISGNGTEHDGASRRHQ-- 225
QY 155 HLQARRGVGPGSDPTSCRGV-----ASGNPAILRALDPYKPRAGFGESESR 204
Db 226 HLVRKFDPMHFKSFDHLNSTSVNETTVAWQDQWEDVIER-----KARS 272
QY 205 RRSRAKRFVSIPIRVETLVVADESVMVKFHGADLEHVLTLTATAARLYRHPSTILNPI 264
Db 273 RAAN-----SWDHVEVLVVDATKMYEHGHSLEDIVLTLSVASTYRHSQSLASIN 326
QY 265 VVYKVLRLDRSDGPKVTGNALTLRNFCAWQKLNKVKSDKHPYWDTAILTRQDLGCA 324
Db 327 VVYKVLRLDRSDGPKVTGNALTLRNFCAWQKLNKVKSDKHPYWDTAILTRQDLGCA 324
QY 325 T-TCDTLGMADVGTMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVNVKVEE 380
Db 387 QGCKDTLGLAELGTMCDMOKSCAIEDNGLSAAFTIAHELGHVFSIPHDDERKSTYMPV 446
QY 381 -----FCKLRAN---HMSPTLIQIDRANPWSACSAAIITDFDLSHG--DCLLD 425
Db 447 NKVKFQSTKEDKTOFQNNFHIMAPTLEYNTHPWSWSPCSAGMLERLENNRGOTQCLEF 506
QY 426 QPSKIPISLPE---DLPGASYTLSSOCELAFGVGSKPCPYMOYCTKLWCTGKAKQWVC 481
Db 507 QPVERRYEDVVRDEPGKKYDAHQCKFVFGPASELCPTMPTCRLMLCATFYQSWMGCR 566
QY 482 TRHFPWADTSCGEGK---LCLKGCACVVRHNLNHRVDGWSNAKWDYPGCSRTCGGGVOLA 539
Db 567 TQHNFPWADTSCDESRSFCHGACVRLAPESLTKIDQWGDWSRSGCSRTCGGGVQKG 626
QY 540 RQCTNPTPANGKYCGVRVYKYSNLEPCPSSASGKSFREQ 583
Db 627 LRDCSPKPRNGKYCVGQGRYRSCNTQECPWDT--QPYREVQ 668

RESULT 3
Q9W493 ID Q9W493 PRELIMINARY; PRT; 1054 AA.
AC Q9W493;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CG4096 protein.
GN CG4096.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
```

Qy	202	RSRRSRGAKRFVSI	PRYVETLVVVADESMVKFGADLEHYLLTLTATAARLYRHPISLNP	261				
Db	309	PDHAEIPRRRSIS	SPRHHVETLIIVADATMSAFH--RDLNGYLLTIMNMVSAKYKDPDGSGNS	367				
Qy	262	INIVVVKLLLRD	RDSGPK--VTGNAALTNRNCAWQKLNKYSDKHPEYWDTAIFLTRQ	319				
Db	368	IEIVVVRLLIQD	EEESQLNLNTONAKNLDRCFSWQHKLNKSEKDPHHHDVAILLTRK	427				
Qy	320	DLGATTCDDTLG	MAVDVTCMDPKRSCSVIEDDGLPSAFTTAAHELGHVFNPHDPNVKVC	378				
Db	428	NIC--ANNCMTL	GLANVGMCKPKQSCSVNEDNGIMLSHTTITHELGHNFGMFHTAKIGCH	486				
Qy	379	EVFGKLRANHM	SPTL-----IQIDRANPWSACSAAITDFDLSGHGDCLLDQPSKPI---	431				
Db	487	PRVGPI--VHTI	PTTFGADTLQV---CWGNCRSKTYITHELDDQGLGEC-LDDPTTLDEY	539				
Qy	432	SLPEDLPGASV	TLSSQCELAFG-----VGSKPCPYMOCYCTKLWCTGKAKGOMVCOTRHF	485				
Db	540	NYTGELPGMRY	NARGCQLQFNLTITDEVGACASAPH-EFCSTLWC--KYNGE--CVTHMR	594				
Qy	486	PWADTSCGEGK	LKLGACVERHNLNKHRYDGVSAKWDPYGPCSCRTGGCVQLARRQCTN	545				
Db	595	PTAPTCLGRNK	WCQKCYRREEL--AAVNGWGMDSEWSECSRSCGGGVSTQORECDN	652				
Qy	546	PTPANGKYCEG	VRVYKRSNLEPCPSSASGKSFREQ	583				
Db	653	PVPANGGVFCI	GERKRYKICKRKFCP--AEEPSFRAQQ	688				
RESULT 4								
Q8TE56	ID	Q8TE56	PRELIMINARY; PRT; 1095 AA.					
AC	Q8TE56;							
DT	01-JUN-2002	(TrEMBLrel. 21, Created)						
DT	01-JUN-2002	(TrEMBLrel. 21, Last sequence update)						
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)						
DE	Metalloprotease disintegrin 17, with thrombospondin domains.							
GN	ADAMTS17.							
OS	Homo sapiens (Human).							
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
OX	NCBI_TaxID=9606;							
EN	[1]							
RP	SEQUENCE FROM N.A.							
RA	MEDLINE=21856482; PubMed=11867212;							
RX	Cal S.; Obaya A.J.; Llamazares M.; Garabaya C.; Quesada V.,							
RA	Lopez-Otin C.;							
RT	"Cloning, expression analysis, and structural characterization of							
RT	seven novel human ADAMTSs, a family of metalloproteinases with							
RT	disintegrin and thrombospondin-1 domains.";							
RL	Gene 283:49-62(2002).							
RL	EMBL; AJ315735; CAC86016.1; -.							
KW	Integrin; Protease.							
DW								
SQ	SEQUENCE 1095 AA; 121099 MW; A5C6540484754D5D CRC64;							
Query Match 25.4%; Score 803.5; DB 4; Length 1095;								
Best Local Similarity 33.5%; Pred. No. 6.2e-58;								
Matches 212; Conservative 93; Mismatches 243; Indels 85; Gaps								
Qy	2	LLLGILTLIAF----	AGRTAGGFEPERVWVPIRLDPDI-----NGRRYYWRGPE	46				
Db	10	LVLPLVLLLVWGL	DPGTAVGDAADVEVLVFWVRPDDVHLPLPAAPGPRRRRPTTP	69				
Qy	47	DS-----GDQGLIFQITAF	EDFYVHLHTPDQAFLPAFSTEHLGVLPVQGTGGSSDLRRC	101				
Db	70	AAPRAPGERALLHL	HPAFGRDLYQLRRLRLRFLSRGFEVEEAGAARR--RGRPAEL--C	125				
Qy	102	FYSGDVNAEPDSF	AAVSLCGGLRCAGF--YGAAYV--ISPLPNASAPAAQRNSQGAHILQ	158				
Db	126	FYSGRVLGHGPSV	LSLSCAAGGLVGLIQLGQEQVLIQLNNSGQPFSGRE-----HLIR	181				
Qy	159	RR-GVPGPSGD---	PTSRCGVASGWNPAILLRALDYPKPRRAGFE--SRSRRSRSGRAKRF	213				

[illegible]

252	Qy	LYRHPSLNPIINVVVVLLLRDRDSQPKVTGNAALTLRNFCAWQKL-NKVSQKHPEYW	310
327	Db	LFKDGTGSDINVVVSLILLEQEPGGILLINHADQSLNFSQWQSQALIKNGKRH----	382
311	Qy	DTAILFRQDLGG--ATTCDTLGMADVGTMCDDPKRSCSVIEDDGLPSAFVTAHELGHVFN	368

Db	556	ICHMLGLAELGTVC\$-SSSCSIVQDTGLPTAFTWAHELGHILNNHDDDDKCMPIYVTRQN	614
Qy	386	AN---HMMSPTLIQIDRANPWSACSAALIITDFLDGHCDCLLDQSPKSI\$P---EDLPG	439
Db	615	NNKVLHIMSSVYGMHPWWSKSRHFVSFELEKTDK\$CL--ET\$VGAHIPYGT\$RLPG	672
Qy	440	ASYT\$SOQCELA\$FVG\$KPCPYMYCTKLMC---TGAKGMQV\$CQTRH\$F\$WADT\$CG-E	495
Db	673	EIY\$LD\$AQCL\$F\$G\$D\$FCY\$CPT\$DECK\$RL\$CNRT\$SGNS\$EQ--C\$ASSNLP\$WADT\$PCG\$SS	730
Qy	496	GKLC\$KLGACV\$ERHNLNKH---RYDGG\$W\$KWD\$PYG\$CSRT\$CGG\$VQL\$ARQ\$CIN\$PT\$PANG	551
Db	731	GHWCQR\$KCV\$---NKHGYRQV\$GW\$G\$W\$PT\$PT\$C\$LL\$CGG\$V\$Q\$BSR\$ECN\$Q\$P\$PENG	786
Qy	552	GKYCEGV\$VKYR\$CN\$LEP\$C\$P\$SS\$AS\$GK\$F\$REEQ	583
Db	787	KYCTGTR\$K\$K\$YR\$SC\$NTHOC\$P\$G\$MDP--REO	816

```

RESULT 8
Q8TE57
ID Q8TE57 PRELIMINARY; PRT; 1072 AA.
AC Q8TE57;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE DE Metalloproteinase disintegrin 16 with thrombospondin type I motif.
GN ADAMTS16.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTS, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains."
RL Gene 283:49-62(2002).
DR EMBL; AJ315734; CAC86015.1; -.
KW Integrin; Protease.
SQ SEQUENCE 1072 AA; 119491 MW; 081E5EFF78F47D061 CRC64;

```

Best Local Similarity 34.0%; Pred. No. 1.7e-54;					
Matches 200; Conservative 74; Mismatches 227; Indels 88; Gaps					
<hr/>					
QY	51	QG LFIQTAFQEDFYHLHTPDAQFLAPAFSTEHLGVPLQGLTGSS-----DLRRCFY 103			
Db	95	ESLHRLKGRPHDFMDLTRSSSIVAPGFVQTL----GKTCTKSQVLTPPEDF--CFY 147	:	:	:
<hr/>					
QY	104	SGDVNAEPDSEFAAVSLCGGLRGAFGYRGAHYVTSPLPNASAPAAQRNSQ---AHLQRR 160	:	:	:
Db	148	QGSLSHRSHNSVALSTCOQLSGMIRTEADYFLRPLPSHLWKLGRAAACGSSPSHVLYKR 207	:	:	:
<hr/>					
QY	161	G---VPGGSGDPISR-----CGVASGNWA-----ILRA 187	:	:	:
Db	208	STEPHAGASELVITSRTWELAHOPLHSSDLRLGLPKQHFCGRKKRYKMPQPPKEDLFIL 267	:	:	:
<hr/>					
QY	188	LDPYKPRRAGFGESRRRSRRGRAKRFVSIPRYVETLVVADESVMVKFHG+ADLEHYLLTLL 246	:	:	:
Db	268	PDEYK---SSLRKHSLLRSRNEEL-----NVETLVVDKKMQNHGHENITYVLTIL 319	:	:	:
<hr/>					
QY	247	ATAARLYRHPSILNPINIVVVKVLLLRDRDSGPKVGTNAAALTLRNFCAWKKL-NKVSOK 305	:	:	:
Db	320	NMYSALFKDGTIGGINIAIVGIIILEDEQPGLVISHHADHTLSSFCQWSQGLMGKDGR 379	:	:	:
<hr/>					
QY	306	HPEYWDTAILFTRODLGG--ATTCDTLLGMADVMTDKPRKSVIEDDDGLPSAFTTAHEL 363	:	:	:
Db	380	H---DHAILLTGLDICKSWKNPECDTLGPAPISGMSKYRSCTINEDTGLGLAFTIAHES 435	:	:	:

QY 364 GHVFNNPHDNV-KVCEEVFGKLRANHMSPPTLIQIDRANPNWSACSAALITFLDSGHGDC 422
 DB 436 GHNFGMIHDEGNCKKSG-----NIMSPTLAGNNGVFSNPSQRSQYLFLSLTAQAIC 490
 QY 423 LLDQPSKPI---SLPDLPGASYTILSQOCELAFVGSXPY---MOYCTKLWC--TGCA 474
 DB 491 LADQP-KPVKEYKYPEKLFGELYDANTOCKWQFGEKAKLMDLDFKDKICKALMCHRIGRK 549
 QY 475 KGMVCOTRHFPAWDTSGEGKCLKGACVERHNLNHRVDGSKWAKWDPYGPCSRTCGG 534
 DB 550 -----CETRFMPAAEGTICGDMWRCGGCVKYGDEGPKPHGHWSNWSNPSRTCGG 604
 QY 535 GVLARQCTNPTNPPANGKYCEGVRYKRSNLEPCPSASGKSFPREQ 583
 DB 605 GVSHRSLCTNPKPSHGKFCBEGSTRTKLCNSQKCPRDS--VDFRAAQ 651

RESULT 9

Q9VF61 PRELIMINARY; PRT; 1229 AA.
 AC Q9VF61;
 DT 01-MAY-2000 (TREMELrel. 13, Created)
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
 DE CG6107 protein.
 GN CG6107.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SOURCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132.
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Balgwin D.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Balgwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houch J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).

EMBL: AE003709; AAF55199.1; --
 FlyBase: FBgn0038340; CG6107.
 DR InterPro: IPR002870; Rep_M12B_propep.
 DR InterPro: IPR001590; Repolysin.
 DR InterPro: IPR000884; TSPI.
 DR InterPro: IPR000130; Zn_MTpeptdse.
 DR Pfam: PF01562; Rep_M12B_propep; 1.
 DR Pfam: PF01421; Repolysin; 1.
 DR Pfam: PF00090; tsp_1; 2.
 DR SMART: SM00209; TSPI; 3.
 DR PROSITE: PS0215; ADAM_MEPRO; 2.
 DR PROSITE: PS0092; TSPI; 2.
 DR PROSITE: PS0142; ZINC_PROTEASE; 1.
 DR Hydrolase; Metalloprotease; Zinc.
 KW SEQUENCE 1229 AA; 138136 MW; C1293180AB19B700 CRC64;

Query Match 23.9%; Score 757.5; DB 5; Length 1229;
 Best Local Similarity 33.3%; Pred. No. 4.9e-54;
 Matches 191; Conservative 91; Mismatches 184; Indels 107; Gaps 27;

QY 55 FOITAFQEDFYHLHTPDPAFLAPAFSTEHLGVPLQG-----LTGGSSDLRR--CFYSGD 106
 DB 195 YNLNVFGRQLHLVLRQDASFV-HNHSMTIRILKEGHEHPGTEAEAEQRLHLCFYSY 253
 QY 107 VNAEPDSFAAIVSLCGGLRGAFYGAEVYISPLPNASAP-----AAQNSQGAHLQ 158
 DB 254 VEDDPHSMVSVSLCGMTGIYIKTSFGALLIQPVNRTSSDEVLRHVRKRSQNRN--HAVS 311
 QY 159 RRGVGGSGDPTSRGCVASGNPAILRALDPYKPRAGFGE-----SRSRRSRAKRFV 214
 DB 312 KFEL-----GLDFMSKLEQVQEEQKSKRLNKKRHYA 347
 QY 215 SIPRVV---ETLVVADESVMKFGHADLEHLLTLTALTAARLYRHPSILNPNIVVVKVL 271
 DB 348 DVDNQVYTLVLIADVNSMKQFHGDIQPYILILMSIVSSIFADASGNSRILLRLIS 407
 QY 272 LRDRDSGPKV---TGNAALTLRNCAMOKLNKYSDKHPEYWDYAILFTRODLGAT--- 325
 DB 408 L-----PNINDOTHSSNMLKHCQF--INQSYER---DTAMLTREPGISVPGK 454
 QY 326 TCDTIGMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNPHDNVKNVCEEVFGKLR 385
 DB 455 ICHMLGLAEGLTVCS--SSSCSIVQDTGLPTAFTWAHELGHILNMNHHDDDDCKMPYVTRQN 513
 QY 386 AN---HMSPTL-IOIDRANPWSACSAALITDLDSDGHDGCLDQPSKPISLP---EDLP 438
 DB 514 NNKVLIHSSVNGIHM---HPMSWSKK---TD-----KSCL--ETSVGAHIPVGTREL 559
 QY 439 GASVTLSSQOCELAFAVGSKPCPYMYQYCTKLWC---TGKAGOMVCQTRHFPWADTSCG- 494
 DB 560 GEIYSLDAQCQLSFGNDFCYCPTDECKRLMCNRTSGNSNEQ--CASNLPWADGTPGCS 617
 QY 495 EGKCLKLGACVERHNLKH---RVDSWAKWDYPCSRCTCGGQVQLARQCTNPTAN 550
 DB 618 SGHNCORGCVS-----NRHGYGRQVNGMGWPTPTPCLSLTCGGGQVQESRECNQPVEN 673
 QY 551 GGYCEGVGVKRYRSNLEPCPSASGKSFPREQ 583
 DB 674 GGYCTGSRKKYRSNTHOCPPGSMDDP--REQQ 704

RESULT 10

Q8TE59 PRELIMINARY; PRT; 1207 AA.
 AC Q8TE59;
 DT 01-JUN-2002 (TREMELrel. 21, Created)
 DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
 DE ADAMTS-19.
 GN ADAMTS19.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

DE	A disintegrin-like and metalloprotease with thrombospondin type 1 motif 14 precursor.
GN	ADAMTS14.
EN	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID:	9606;
RN	[1]
SEQUENCE FROM N.A.	
RP	MEDLINE-21638061; PubMed=11779638;
RX	Bolz H., Ramirez A., von Bredelow B., Kubisch C.;
RT	"Characterization of ADAMTS14, a novel member of the ADAMTS metalloproteinase family.";
RL	Biochim Biophys Acta 1522:221-225(2001).
DR	EMBL; AF358666; AAL40229.1;
DR	InterPro: IPR002870; Pep_M12B_propep.
DR	InterPro: IPR001590; Reprolysin.
DR	InterPro: IPR000884; TSPI.
DR	Pfam; PF01562; Pep_M12B_propep; 1.
DR	Pfam; PF01421; Reprolysin; 1.
DR	Pfam; PF00090; tsp_1; 4.
DR	SMART; SM00209; TSPI; 4.
DR	PROSITE; PS00215; ADAM_MEPRO; 1.
DR	PROSITE; PS0092; TSPI; 1.
KW	SIGNAL; Integrin; Protease; Metalloprotease.
FT	SIGNAL 1 22 POTENTIAL.
SEQ	SEQUENCE 1223 AA; 133871 MW; 36394ACAD92F170F CRC64;
Query Match	22.6%; Score 715.5; DB 4; Length 1223;
Best Local Similarity	31.2%; Pred. No. 1.5e-50;
Matches	175; Conservative 86; Mismatches 238; Indels 61; Gaps
QY	45 PEDSGDGLIFQTAFQEDFYHLTDPAQFLAPAFSTEH-----LGVPLOGLTGSSD 97
DB	98 PGRVGRHSLYENVTVFGKELHLRLRNRLIVPGSSVEHQEDREFLRQLR----- 149
QY	98 LRCFTYSGDVNABPDGFAAVSLCGGLRGAFGYRGAEYVISPLPNASAPAAQRNSOG-AHL 156
DB	150 -QECVTGGTGTMPGAVALISNGDLAGLTDTSTDFIEPLERGO--QEKEASGRTHV 205
QY	157 LORRGVP---GGPSGDPTSRGCVASGWNPAILRALDPYKPRAGGFSRRSRRSGRAKR 212
DB	206 VYREAVQQEWAPDGDDLNE-AFGLDGLPULLGLVGDO-----LGDTERKKRAKPGS 258
QY	213 FVSIPRYVETLVVADESWMKFHAD-LEHYLLTLATAARLYRHPSILNPINIVVKVLL 271
DB	259 Y-----SIEVLVVDDSVFRGHKVONTVLTLMNIDEIYHDESIGLVHINIALVRLLM 313
QY	272 LRORDSPKV-TGNAALTLENFCAWOKKLNVSKDKHPKYWDTAILETRDOLCGATTCDFL 330
DB	314 VGYSLSLTERGNPSRSLEQVCRAWHSOQRQDPSHAHHHDHVVFITRODF-----GPS 367
QY	331 GMADVGMCDPKRCSVIEDDGLPSAFTTAHELGHVFNPHD---NVKVCCEEVFGKLRAN 387
DB	368 GIAPVTGMCHPLRSCALNHEDGESSAFVIAHETGVHLMGHEHGQGNCADETSLSGS---- 423
QY	388 HMSPPTLIQIDRANPWSACSAAIITFDLSGHGDCILLDQSPKI-SLPEDLPGASVTLSQ 446
DB	424 -VMAPLVQAFAHRFHWRSCSKLELSRYLPS--YDCLDDPDFPAWPQPPELPGINYSMD 480
QY	447 OCELAFGVGSKP---PYMOYCTKLACTGAKGQMVCOTRHPFWADGTCGEGEKLCCLKCA 503
DB	481 QCREFDGGSYOTCUAFRTTFEPCKQLWCS-HPDNPFCKTKKGPPLOGTECAPGKCWKFGH 539
QY	504 CVERHNLNKHVDGSWAkWDPYPGCSRCTGGGVLARRQCTNPTPANGKCYGEGVRKYR 563
DB	540 CIWKSPEQTYQDGGWSSWTIKFGSCSRSCGGVRSRSCNNPSPAYGGRPCLGPMFEYQ 599
QY	564 SCNLPECPSSASGKSFREEQ 583
DB	600 VCNSEECPGTY--EDFRAQQ 617

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:25 ; Search time 28.3707 Seconds
(without alignments)
2738.218 Million cell updates/sec

Title: US-10-009-332-1_COPY_1_583

Perfect score: 3163

Sequence: 1 MLLGILTLFAGTAGGFE.....SCNLEPCPSSAGSKSFRERQ 583

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3163	100.0	950	22	AAG62299 Human metalloprotease
2	3155	99.7	928	23	AAU72899 Human metalloprotease #2
3	3155	99.7	950	23	AAE22541 Human protease PRT
4	3155	99.7	952	23	AAU74751 Rat metalloprotease #3
5	2497.5	79.0	505	21	AAE221257 Human metalloprotease #1
6	1913	60.5	367	23	AAE22542 Human protease #1
7	1767	55.9	381	21	AAE221261 Human protease #1
8	1689	52.8	321	23	AAE22540 Mouse metalloprotease
9	1562.5	49.4	896	21	AAE221265 Human METH1 protei
10	1560.5	49.3	950	20	AAU49501

11	1560.5	49.3	950	22	AAB73549 Human ADAM-type me
12	1560.5	49.3	950	22	AAB50002 Human METH1. Homo
13	1560.5	49.3	967	19	AAW80285 Human integrin lig
14	1560.5	49.3	968	22	AAB50011 Protein; SEQ ID 12
15	1559.5	49.3	967	20	AAU04142 Human Taogo-71 pro
16	1538.5	48.6	967	20	AAW78189 Human secreted pro
17	1536	48.6	950	21	AAU53899 Amino acid sequenc
18	1362	43.1	890	20	AAU49502 Human METH2 protei
19	1362	43.1	890	22	AAB50003 Human METH2. Homo
20	1359	43.0	889	22	AAU74946 Human ADAM type me
21	1351	42.7	727	20	AAW78435 Human ADAMTS-1 pro
22	1338.5	42.3	837	20	AAW75425 Human aggrecan deg
23	1337.5	42.3	837	21	AAU99429 Human PRO1563 (UNQ
24	1337.5	42.3	837	22	AAU29199 Human PRO polypept
25	1337.5	42.3	837	22	AAB66178 Protein of the inv
26	1336.5	42.3	837	22	AAU78228 Human aggrecanase-
27	1336.5	42.3	840	21	AAB21256 Human metalloprote
28	1329.5	42.0	905	22	Murine ADAMTS-8 am
29	1240	39.2	930	22	AAU72284 Human ADAMTS-5 am
30	1232.5	39.0	947	22	AAU86950 Human metalloprote
31	1232.5	39.0	1629	23	ABG30703 Human aggrecanase
32	1232.5	39.0	1629	23	AAO14448 Human ADAMTS-SI pr
33	1232.5	39.0	1916	23	AAE19173 Human protease, PR
34	1232.5	39.0	1935	23	AAU72896 Human metalloprote
35	1229.5	38.9	1602	23	ABG30702 Human aggrecanase
36	1226.5	38.8	1073	21	AAU21284 Human metalloprote
37	1217.5	38.5	870	21	AAU21252 Rat metalloprote
38	1212.5	38.3	930	20	AAU75426 Human aggrecan deg
39	1212	38.3	1934	22	AAU72301 Human ADAMTS-9 alt
40	1205	38.1	929	21	AAU41226 Human ORFX ORF990
41	1160	36.7	1505	23	AAU72897 Human metalloprote
42	1153.5	36.5	625	23	AAU48394 Rat aggrecanase; #12
43	1149.5	36.3	1907	23	AAU77133 Human protease #12
44	1142.5	36.1	680	21	AAU21251 Human metalloprote
45	1121	35.4	1882	22	AAU72286 Human ADAMTS-9 am

ALIGNMENTS

RESULT 1

AAG62299

ID AAG62299 standard; protein; 950 AA.

XX AAG62299;

AC AAG62299;

XX 23-AUG-2001 (first entry)

DT Human metalloprotease MDT56 protein.

DE Metalloprotease; human; aggrecanase; joint disease; osteoarthritis;

KW osteopathic; antiarthritic.

XX Homo sapiens.

XX WO200134785-A1.

XX 17-MAY-2001.

XX 10-NOV-2000; 2000WO-JP07917.

XX 11-NOV-1999; 99JP-0321740.

XX 16-MAY-2000; 2000JP-0144020.

XX (YAMA) YAMANOUCHI PHARM CO LTD.

XX (KAZU-) KAZUSA DNA RES INST.

XX Yamaaji N, Nishimura K, Abe K, Ohara O, Nagase T, Nomura N;

XX WPI; 2001-343602/36.

XX N-PSDB; AA41003.

PT Metalloprotease with aggrecanase activity for treating joint diseases

KW	anorectic; antiinflammatory; aspartyl protease; cysteine protease;
KW	metalloprotease; serine protease; cancer; haematopoietic; breast;
KW	lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;
KW	immune-related disease; cardiovascular disease; neuronal disease;
KW	migraine; sexual dysfunction; mood disorder; attention disorder;
KW	cognition disorder; hypotension; hypertension; psychotic disorder;
KW	dyskinesia; metabolic disorder; inflammatory disorder.
XX	
OS	Homo sapiens.
PN	WO200103782-A2.
PD	08-NOV-2001.
PF	04-MAY-2001; 2001WO-US14431.
PR	04-MAY-2000; 2000US-201879P.
PA	(SUGEN-) SUGEN INC.
PI	Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
PI	Payne V;
XX	
DR	WPI; 2002-041502/05.
DR	N-PSDB; AAS97182.
XX	
PT	Novel protease polypeptide useful for screening for substances that may
PT	be used to treat, e.g., cancers, immune-related diseases,
PT	cardiovascular disease, migraine, pain, psychotic and inflammatory
PT	disorders -
PS	Claim 28; Figure 2G; 232pp; English.
XX	
CC	The invention relates to an isolated, enriched, or purified protease
CC	polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
CC	screen for substances (S) that may modulate its activity. Administering
CC	S (which modulates protease activity in vitro) may be used to treat a
CC	disease or disorder selected from cancers (e.g., of tissues, of blood or
CC	haematopoietic origin, of the breast, colon, lung, prostate, cervical,
CC	brain, ovarian, bladder or kidney), immune-related diseases and
CC	disorders, cardiovascular disease, brain or neuronal-associated diseases
CC	(e.g., central or peripheral nervous system diseases, migraine, pain,
CC	sexual dysfunction, mood disorders, attention disorders, cognition
CC	disorders, hypotension, hypertension, psychotic disorders, neurological
CC	disorders and dyskinesias), metabolic disorders and inflammatory
CC	disorders. (I) may also be useful as a diagnostic tool for a disease or
CC	disorder such as those above. AAU72876-AAU72910 represent human
CC	protease amino acid sequences of the invention.
XX	
QQ	Sequence 928 AA;
Query Match	99.7%; Score 3155; DB 23; Length 928;
Best Local Similarity	99.8%; Pred. No. 3.3e-254;
Matches 582; Conservative	0; Mismatches 1; Indels 0; Gaps
QY	1 MLLGIILTFAGRTAGGFEPEVVPVPIRLDPDINGRRYYWRGPDSDGGLFIQITAF 60
DB	26 MLLGIILTFAGRTAGGFEPEVVPVPIRLDPDINGRRYYWRGPDSDGGLFIQITAF 85
QY	61 QEDPYLHUTDPAQFLAPAFSTEHGLVPLQGLTGSSDLRCFCFSGDVNAEPDFAAVSLC 120
DB	86 QEDPYLHUTDPAQFLAPAFSTEHGLVPLQGLTGSSDLRCFCFSGDVNAEPDFAAVSLC 145
QY	121 GGLRGAFYRGAEVVISPLPNASAPAAQRNSQGAHLQRRVPGGPGSDPTSRCGVASGW 180
DB	146 GGLRGAFYRGAEVVISPLPNASAPAAQRNSQGAHLQRRVPGGPGSDPTSRCGVASGW 205
QY	181 NPAILRALDPYKPRRAGFSGESRRRRSGRAKRFVSIPIRYVETLTVVADESMVKFHGADLEH 240
DB	206 NPAILRALDPYKPRRAGFSGESRRRRSGRAKRFVSIPIRYVETLTVVADESMVKFHGADLEH 265
QY	241 YLLPLLTAARLYRHPSSLINPINIVVYKLLLRDSDSGPKVTGNAALTLRNFCAWOKKLN 300

Db 266 YLLTLLATAARLYRHPSILNPINIVVKKVLLLRDRDSGPKVTGNALTLRNFCWQKKLN 325
 QY 301 KVSDEHPEYWDTAILFTRODLGGATCTDLGMADVGMCDPKRSCSVIEDDGLPSAFTTA 360
 Db 326 KVSDEHPEYWDTAILFTRODLGGATCTDLGMADVGMCDPKRSCSVIEDDGLPSAFTTA 385
 QY 361 HELGHVFNPHDNVNVCEVEFGKLRANHMSPTLIQIDRANPWSACSAAIITDFLDGSHG 420
 Db 386 HELGHVFNPHDNVNVCEVEFGKLRANHMSPTLIQIDRANPWSACSAAIITDFLDGSHG 445
 QY 421 DCLLDQPSKPISLPEDLPASVYLSQOCELAFGVGSKPCPYMOYCTKLWCTGKAGQMVC 480
 Db 446 DCLLDQPSKPISLPEDLPASVYLSQOCELAFGVGSKPCPYMOYCTKLWCTGKAGQMVC 505
 QY 481 QTRHFPWADGTSCEGKLCGLKACVERHNLNKHRVDGSAKWADPYGPCSRTCGGGVQLAR 540
 Db 506 QTRHFPWADGTSCEGKLCGLKACVERHNLNKHRVDGSAKWADPYGPCSRTCGGGVQLAR 565
 QY 541 RQCTNPTPANGKYCEGVYRVKRSNLEPCPSSASGKSFREQ 583
 Db 566 RQCTNPTPANGKYCEGVYRVKRSNLEPCPSSASGKSFREQ 608

RESULT 3
 AAE22541
 ID AAE22541 standard; Protein; 950 AA.

XX AAE22541;
 XX 26-JUL-2002 (first entry)
 XX Human protease #2.

XX Human; novel human protein; NHP; protease; biological disorder; obesity;
 KW high blood pressure; arthritis; connective tissue disorder; infertility;
 KW gene therapy; enzyme.
 XX Homo sapiens.

XX WO200226949-A2.
 XX 04-APR-2002.
 XX 27-SEP-2001; 2001WO-US30350.
 XX 29-SEP-2000; 2000US-236689P.
 XX (LEXI-) LEXICON GENETICS INC.

XX Friddle CJ, Hilbun E;
 XX WPI; 2002-372123/40.
 DR N-PSDB; AAD35569.
 XX Novel nucleic acid encoding a human protease, useful as a hybridization
 PT probe for screening libraries and assessing gene expression patterns -
 PS Claim 6; Page 36-38; 41pp; English.

XX The present sequence is novel human protein (NHP), human protease.
 CC NHPs share structural similarity with animal proteases particularly
 CC zinc metalloproteases. Sequences of the invention are useful in
 CC therapeutic, diagnostic and pharmacogenomic applications. NHP
 CC polynucleotides are used as hybridisation probes for screening
 CC libraries and assessing gene expression patterns. They can also be
 CC used for treating related biological disorders such as obesity, high
 CC blood pressure, arthritis, connective tissue disorders and infertility.
 CC They are also used in gene therapy.

XX Sequence 950 AA;

Query Match 99.7%; Score 3155; DB 23; Length 950;
 Best Local Similarity 99.8%; Pred. No. 3.4e-254.

Matches 582; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MLLLGILTLAFAGRTAGGPEPEVVPVIRLPDINGRRYYWRGPDSDGQGLIFOITAF 60
 Db 1 MLLLGILTLAFAGRTAGGPEPEVVPVIRLPDINGRRYYWRGPDSDGQGLIFOITAF 60
 QY 61 QEDFYHLHPDPAQFLAPAFSTEHLGVPLOGLTGGSSDLRRRCFYSGDVNAEPDSFAAVSLC 120
 Db 61 QEDFYHLHPDPAQFLAPAFSTEHLGVPLOGLTGGSSDLRRRCFYSGDVNAEPDSFAAVSLC 120
 QY 121 GGLRGAFGVRGAYVLSPLNASAPAAQRNSQGAHLQLRRGVPGGPGSDPTSRGCVASGW 180
 Db 121 GGLRGAFGVRGAYVLSPLNASAPAAQRNSQGAHLQLRRGVPGGPGSDPTSRGCVASGW 180
 QY 181 NPAILRALDPYKPRRAGFGESRRRSRAKRFVSIPIRYVETLVVADESMMVKFHHGADLEH 240
 Db 181 NPAILRALDPYKPRRAGFGESRRRSRAKRFVSIPIRYVETLVVADESMMVKFHHGADLEH 240
 QY 241 YLLTLLATAARLYRHPSILNPINIVVKKVLLLRDRDSGPKVTGNALTLRNFCWQKKLN 300
 Db 241 YLLTLLATAARLYRHPSILNPINIVVKKVLLLRDRDSGPKVTGNALTLRNFCWQKKLN 300
 QY 301 KVSDEHPEYWDTAILFTRODLGGATCTDLGMADVGMCDPKRSCSVIEDDGLPSAFTTA 360
 Db 301 KVSDEHPEYWDTAILFTRODLGGATCTDLGMADVGMCDPKRSCSVIEDDGLPSAFTTA 360
 QY 361 HELGHVFNPHDNVNVCEVEFGKLRANHMSPTLIQIDRANPWSACSAAIITDFLDGSHG 420
 Db 361 HELGHVFNPHDNVNVCEVEFGKLRANHMSPTLIQIDRANPWSACSAAIITDFLDGSHG 420
 QY 421 DCLLDQPSKPISLPEDLPASVYLSQOCELAFGVGSKPCPYMOYCTKLWCTGKAGQMVC 480
 Db 421 DCLLDQPSKPISLPEDLPASVYLSQOCELAFGVGSKPCPYMOYCTKLWCTGKAGQMVC 480
 QY 481 QTRHFPWADGTSCEGKLCGLKACVERHNLNKHRVDGSAKWADPYGPCSRTCGGGVQLAR 540
 Db 481 QTRHFPWADGTSCEGKLCGLKACVERHNLNKHRVDGSAKWADPYGPCSRTCGGGVQLAR 540
 QY 541 RQCTNPTPANGKYCEGVYRVKRSNLEPCPSSASGKSFREQ 583
 Db 541 RQCTNPTPANGKYCEGVYRVKRSNLEPCPSSASGKSFREQ 583

RESULT 4
 AAU74751
 ID AAU74751 standard; Protein; 952 AA.

XX AAU74751;
 XX 09-APR-2002 (first entry)

XX Human protease PRTS-11 protein sequence.

XX Human; protease; PRTS; gastrointestinal; Crohn's disease; cancer;
 KW cardiovascular; atherosclerosis; autoimmune disorder; dermatitis;
 KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS;
 KW cell proliferative disorder; developmental disorder; epilepsy;
 KW Duchenne muscular dystrophy; epithelial disorder; neurological disorder;
 KW reproductive disorder; endometriosis.

XX Homo sapiens.

XX WO200198468-A2.

XX 27-DEC-2001.

XX 13-JUN-2001; 2001WO-US19178.

XX 16-JUN-2000; 2000US-212336P.

XX 22-JUN-2000; 2000US-213955P.

XX 29-JUN-2000; 2000US-21396P.

XX 07-JUL-2000; 2000US-216821P.

XX 14-JUL-2000; 2000US-218946P.

XX (INCY-) INCYTE GENOMICS INC.
 XX Yue H, Elliott VS, Gandhi AR, Lal P, Au-young J, Tribouley CM;
 PI Deleage AM, Baughn MR, Nguyen DB, Lee EA, Hafalia A, Khan FA;
 PI Wallia NK, Yao MG, Lu DAM, Patterson C, Tang YT, Walsh RT;
 PI Azimzai Y, Lu Y, Ramkumar J, Xu Y, Reddy R, Das D, Kearney L;
 PI Kalllick DA;
 XX WPI: 2002-090437/12.
 DR N-PSDB; ABK12894.
 XX
 XX Twenty one human proteases (referred to as PRS-1 to PRS-21), useful
 PT in the diagnosis, treatment and prevention of gastrointestinal (e.g.
 PT gastritis), cardiovascular (e.g. atherosclerosis) and cell
 PT proliferative (e.g. cancer) disorders -
 XX
 XX Claim 1; Page 144-146; 177pp; English.
 XX
 XX The present invention relates to twenty one new human proteases,
 CC referred to as PRS-1 to PRS-21. The PRS polynucleotides and
 CC polypeptides of the invention are useful in the diagnosis, treatment and
 CC prevention of gastrointestinal e.g. gastritis, esophageal carcinoma and
 CC Crohn's disease, cardiovascular e.g. atherosclerosis, hypertension and
 CC myocardial infarction, autoimmune/inflammatory e.g. acquired
 CC immunodeficiency syndrome (AIDS), allergies and osteoarthritis, cell
 CC proliferative e.g. cancer, developmental e.g. Duchenne and Becker
 CC muscular dystrophy, epithelial e.g. dermatitis, neurological e.g.
 CC epilepsy and Alzheimer's disease and reproductive e.g. infertility and
 CC endometriosis disorders. Numerous other examples of each disorder are
 CC given in the specification. The present protein sequence represents
 CC the human protease PRS-11 protein of the invention.
 XX
 XX Sequence 952 AA;
 SQ
 Query Match 99.7%; Score 3155; DB 23; Length 952;
 Best Local Similarity 99.8%; Pred. No. 3.4e-254;
 Matches 582; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MLLGILTLAFAGTAGGFPREVVVPIRLDPDINGRRYWRGPDSDGGLIFQITAF 60
 DB 1 MLLGILTLAFAGTAGGFPREVVVPIRLDPDINGRRYWRGPDSDGGLIFQITAF 60
 QY 61 QEDFLHLTPDAQPLAFSTEHLGVPLQGLTGSSDLRCFCYSGDVNAEPDSFAAVSLC 120
 DB 61 QEDFLHLTPDAQPLAFSTEHLGVPLQGLTGSSDLRCFCYSGDVNAEPDSFAAVSLC 120
 QY 121 GGLRGAFYRGAEYVISPPLNAPAAQRNSOGAHLQRRGVPDPSGDPTRSCGVASGW 180
 DB 121 GGLRGAFYRGAEYVISPPLNAPAAQRNSOGAHLQRRGVPDPSGDPTRSCGVASGW 180
 QY 181 NPAILRALDPYKPRRAGFGESRRSRGKRVFVSPRYVETLVADESVMKFGADLEH 240
 DB 181 NPAILRALDPYKPRRAGFGESRRSRGKRVFVSPRYVETLVADESVMKFGADLEH 240
 QY 241 YLLTLLAARLYRHPILNIPINIVVKVLLLRDRSGPKVTGNAALTIRNCAWOKLN 300
 DB 241 YLLTLLAARLYRHPILNIPINIVVKVLLLRDRSGPKVTGNAALTIRNCAWOKLN 300
 QY 301 KVSQKHEPYWDATLFTLRQDLGGATTCDTLGMADVGTMCDPKRSVIEDDGLPSAFTTA 360
 DB 301 KVSQKHEPYWDATLFTLRQDLGGATTCDTLGMADVGTMCDPKRSVIEDDGLPSAFTTA 360
 QY 361 HELGHVFNPHNDNVKVEEYFGKLRANHMSPILQIDRANPWSACSAAITDLDGSHG 420
 DB 361 HELGHVFNPHNDNVKVEEYFGKLRANHMSPILQIDRANPWSACSAAITDLDGSHG 420
 QY 421 DCLLDQPSKPISTLPEDLPAGSYTILSQOCELAFGVSKPCPYMOCYTKLWCTGKAKGMVC 480
 DB 421 DCLLDQPSKPISTLPEDLPAGSYTILSQOCELAFGVSKPCPYMOCYTKLWCTGKAKGMVC 480
 QY 481 QTRHFPNADGTSCEGKGLCLKGACVERHNLNKRVDSWAKWDPYGPCSRTCGGVOLAR 540

DB 481 QTRHFPNADGTSCEGKGLCLKGACVERHNLNKRVDSWAKWDPYGPCSRTCGGVOLAR 540
 QY 541 RQCTNPFPANGKYCEGVVYKRSNLEPCPSSASGKSFREEQ 583
 DB 541 RQCTNPFPANGKYCEGVVYKRSNLEPCPSSASGKSFREEQ 583
 RESULT 5
 AAB21257
 ID AAB21257 standard; Protein; 505 AA.
 XX
 AC AAB21257;
 XX
 DT 23-FEB-2001 (first entry)
 XX
 DE Rat metalloproteinase ADAMTS-5.
 XX
 KW Rat; ADAMTS-5; metalloproteinase; ADAM;
 KW a disintegrin and metalloproteinase domain; thrombospondin domain;
 KW vaccine; neurotropic; neuroprotective; antiparkinsonian;
 KW cerebroprotective; cytoskeletal; antiarthritic; immunosuppressive;
 KW Alzheimer's disease; Parkinson's disease; stroke; cancer; arthritis;
 KW autoimmune disease; brain tumour; brain injury.
 XX
 OS Rattus norvegicus.
 XX
 PN WO200053774-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US062237.
 XX
 PR 08-MAR-1999; 99US-0264585.
 XX
 PA (NEUR-) NEUROCRINE BIOSCIENCES INC.
 XX
 PI Kelner GS, Clark M, Maki RA;
 XX
 DR WPI: 2000-594326/56.
 DR N-PSDB; AAA95827.
 XX
 PT Polynucleotide encoding novel members of a disintegrin,
 PT metalloproteinase and thrombospondin domain protein family used to
 PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
 PS Claim 12; Fig 14; 129pp; English.
 XX
 CC The present sequence is rat metalloproteinase ADAMTS-5. The
 CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
 CC and Metalloproteinase Domain) family. Members of the ADAMTS family
 CC contain a thrombospondin domain in addition to the disintegrin and
 CC metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
 CC useful for the manufacture of medicaments for treating conditions
 CC associated with neuroinflammation and/or neurodegeneration, such as
 CC Alzheimer's disease, Parkinson's disease and stroke. They are also
 CC useful for treating conditions associated with cell proliferation, cell
 CC migration, inflammation and/or angiogenesis, such as cancer, arthritis,
 CC and autoimmune diseases. They can be used to treat patients afflicted
 CC with an invasive tumour, a brain tumour or brain injury.
 XX
 SQ Sequence 505 AA;

Query Match 79.0%; Score 2497.5; DB 21; Length 505;
 Best Local Similarity 93.5%; Pred. No. 1e-199;
 Matches 462; Conservative 8; Mismatches 21; Indels 3; Gaps 2;
 QY 89 QGLTGSSDLRCFCYSGDVNAEPDSFAAVSLCGLRGAFYRGAEYVISPPLNAPAAQ 148
 DB 13 QRLTGSSDLRCFCYSGYVNAEPDSFAAVSLCGLRGAFYRGAEYVISPPLNAPAAQ 72
 QY 149 RNSOGAHLQRRGVPDPSGDPTRSCGVASGWNPAIRLALDPYKPRRAGFGESRRSRG 208
 DB 73 RNSOGAHLQRRGVPDPSGDPTRSCGVASGWNPAIRLALDPYKPRRAGFGESRRSRG 132


```

PR 08-MAR-1999; 99US-0264585.
XX (NEUR-) NEUROCRINE BIOSCIENCES INC.
XX Kelner GS, Clark M, Maki RA;
XX WPI; 2000-594326/56.
DR N-PSDB; AAA95831.
XX Polynucleotide encoding novel members of a disintegrin,
PT metalloproteinase and thrombospondin domain protein family used to
PT prevent and treat Alzheimer's disease, cancer and autoimmune diseases -
XX Claim 12; Fig 23; 129pp; English.
PS
XX The present sequence is human metalloproteinase ADAMTS-5. The
CC ADAMTS family of proteins is closely related to the ADAM (A Disintegrin
CC and Metalloproteinase Domain) family. Members of the ADAMTS family
CC contain a thrombospondin domain in addition to the disintegrin and
CC metalloproteinase domains found in the ADAMs. ADAMTS polypeptides are
CC useful for the manufacture of medicaments for treating conditions
CC associated with neuroinflammation and/or neurodegeneration, such as
CC Alzheimer's disease, Parkinson's disease and stroke. They are also
CC useful for treating conditions associated with cell proliferation, cell
CC migration, inflammation and/or angiogenesis, such as cancer, arthritis
CC and autoimmune diseases. They can be used to treat patients afflicted
CC with an invasive tumour, a brain tumour or brain injury.
XX
XX Sequence 381 AA;
SQ
Query Match 55.9%; Score 1767; DB 21; Length 381;
Best Local Similarity 93.3%; Pred. No. 6.3e-139;
Matches 322; Conservative 3; Mismatches 18; Indels 2; Gaps 2;
QY 240 HYLLTLTAARLYRHPSTLNINIVVVKVLLLRDRDSCPKVTGNAAALTLRNFCAWQKKL 299
DB 2 HYRAAARAG-IFKHPSILNINIVVVKVLLLRDRDSCPKVTGNAAALTLRNFCAWQKKL 60
QY 300 NKVSDKHPEYWDTAIFLTRODLGATTCDTLGMADVTGMDPKRSCSVIEDDGLPSAFTT 359
DB 61 NKVSDKHPEYWDTAIFLTRODLGATTCDTLGMADVTGMDPKRSCSVIEDDGLPSAFTT 120
QY 360 AHELGHVFNPHDNVKEEVFGKLRANHMMSPTLIQIDRANPWSACSAIITDFLDSGH 419
DB 121 AHELGHVFNPHDNVKEEVFGKLRANHMMSPTLIQIDRANPWSACSAIITDFLDSGH 180
QY 420 GDCLLDQPSKPISLPDLPGASVTLTQQCELAFGVSKPCPYMOYCTKLWCTGKAKGMV 479
DB 181 GDCLLDQPSKPIELPDLPGASVTLTQQCELAFGVSKPCPYMOYCTKLWCTGKAKGMV 240
QY 480 CQTRHFPWADTSCGEGKCLKACACVERHNLNKHRYVDGSMKWDPYGPCSRTCGGVQLA 539
DB 241 CQTRHFPWADTSCGEGKCLKACACVERHNLNKHRYVDGSMKWDPYGPCSRTCGGVQLA 300
QY 540 RROQTHPNT-ANGKVCYGVVYKRCNLEPCPSSASGKSFRREQ 583
DB 301 RROQHOPXPLTGGKYCEGVVYKRCNLEPCPSSASGKSFRREQ 345
RESULT 8
AAE22540
ID AAE22540 standard; Protein; 321 AA.
XX
AC AAE22540;
XX
DT 26-JUL-2002 (first entry)
XX
DE Human protease #1.
XX
KW Human; novel human protein; NHP; protease; biological disorder; obesity;
KW high blood pressure; arthritis; connective tissue disorder; infertility;
KW gene therapy; enzyme.
XX

```

```

OS Homo sapiens.
XX WO200226949-A2.
XX 04-APR-2002.
XX 27-SEP-2001; 2001WO-US30350.
XX 29-SEP-2000; 2000US-236689P.
XX (LEXI-) LEXICON GENETICS INC.
XX Friddle CJ, Hilbun E;
XX WPI; 2002-372123/40.
XX N-PSDB; AAD35568.
XX Novel nucleic acid encoding a human protease, useful as a hybridization
XX probe for screening libraries and assessing gene expression patterns -
XX Claim 5; Page 34-35; 41pp; English.
XX The present sequence is novel human protein (NHP), human protease.
XX NHPs share structural similarity with animal proteases particularly
XX zinc metalloproteases. Sequences of the invention are useful in
XX therapeutic, diagnostic and pharmacogenomic applications. NHP
XX polynucleotides are used as hybridisation probes for screening
XX libraries and assessing gene expression patterns. They can also be
XX used for treating related biological disorders such as obesity, high
XX blood pressure, arthritis, connective tissue disorders and infertility.
XX They are also used in gene therapy.
XX
XX Sequence 321 AA;
SQ
Query Match 52.8%; Score 1669; DB 23; Length 321;
Best Local Similarity 99.4%; Pred. No. 7.5e-131;
Matches 318; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYWRGPDSDGGLIFQITAF 60
DB 1 MLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYWRGPDSDGGLIFQITAF 60
QY 61 QEDFYLHTLPAQFLAPAFSTHGLGVPLQGLTGSSDLRRCFYSGDVNAEPDFAAVSLC 120
DB 61 QEDFYLHTLPAQFLAPAFSTHGLGVPLQGLTGSSDLRRCFYSGDVNAEPDFAAVSLC 120
QY 121 GGLRGAFGYRGAEEYVISPPLNASAPAAQNSQGAHLQLQRRGVPGSGDPTSRGCVASGW 180
DB 121 GGLRGAFGYRGAEEYVISPPLNASAPAAQNSQGAHLQLQRRGVPGSGDPTSRGCVASGW 180
QY 181 NPAILRALDPYKPRRAGGESRRSRRAKRFVSIPIRYVETLVVADESVMKFFGADLEH 240
DB 181 NPAILRALDPYKPRRAGGESRRSRRAKRFVSIPIRYVETLVVADESVMKFFGADLEH 240
QY 241 YLLTLLATAARLYRHPSTLNINIVVVKVLLLRDRDSCPKVTGNAAALTLRNFCAWQKKLN 300
DB 241 YLLTLLATAARLYRHPSTLNINIVVVKVLLLRDRDSCPKVTGNAAALTLRNFCAWQKKLN 300
QY 301 KVSDDKHPEYWDTAIFLTREQ 320
DB 301 KVSDDKHPEYWDTAIFLTREQ 320
RESULT 9
AAE21265
ID AAE21265 standard; Protein; 896 AA.
XX
AC AAE21265;
XX
DT 23-FEB-2001 (first entry)
XX
DE Mouse metalloproteinase ADAMTS-1.
XX

```


or mobilisation (chemotaxis) of immune cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or toxins), or infectious. They can also be used to treat inflammatory conditions, both chronic and acute conditions. The products can also be used for detection and diagnosis. AA232002 to AA232080, and AA49503 to AA49511 represent sequences given in the exemplification of the present invention.

XX Sequence 950 AA;

Query Match 49.3%; Score 1560.5; DB 20; Length 950;
Best Local Similarity 51.6%; Pred. No. 3.7e-121;
Matches 320; Conservative 73; Mismatches 160; Indels 67; Gaps 17;

```

QY 1 MLLGLTLTAFAGTAGGPEPEREVVPIRLDPDINGRRYYWRGPEDSQDGLIFQITAF 60
DB 19 LLLAAALLAVSDALGRPEDEELVVP-ELE-----RAP---GHGTRRLRHAF 64
QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRCFYSGDVNAEPDSF 114
DB 65 DQDLLELRPDSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCFYSGTVNGDPSSA 119
QY 115 AAVSLCGGLRGAFYGAAYVISPPLNAS---APAAQNSQGA---HLLQ---RRGVPGG 165
DB 120 AALSICGVRGAFYILGEAYFIQPLPAASERLATAAPEKPPAPLQFHLRLRNROGDVGG 179
QY 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPYKPRAGFGESESRRS 207
DB 180 TCGVVDDEPRPTGKARTDEDEGEDEGPQWS-----PQDPALQGVGP-TGTGS 230
QY 208 GRAKRFVSPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPISILNPINVVV 267
DB 231 IRKKRFVSHRYVETLMVADQSMAPFGSLKHYLLTLFSAARLYKHPISIRNSVLVV 290
QY 268 KVLILLDRDSDGPKVTGNAAALTLRNFCAWOKLKNKYSDKHPEYWDTAILFTRODLCGATC.327
DB 291 KILVIHDEQKPEVTSNAALTLRNFCNWKQHNPPSDRDAEHYDTAILFTRODLCGQTC 350
QY 328 DTGLMADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVFNPHDNKVCVEEVFGKLRAN 387
DB 351 DTGLMADVGTCDPSRSCSVIEDDGLQAAFTTAHELGHVFNPHDDAKQASLNGVQDS 410
QY 388 HMASPTLIQIDRANPWSACSAIITDFLDSGHGDCLLDQSPKPSISLPEDLPASVTL 447
DB 411 HMASMLNLDHSQSPWSPCSAYMITSFLDNGHGECLMDKPNQIQLPGDLPGTSDANRQ 470
QY 448 CELAFGVGSKPCP-YMQYCTKLWCTGKAKGOMVQCTRHFPWADGTSCGEGKLCAGACVE 506
DB 471 CQFTFGEDSKHCPDAASTCTLMCTSGVLVCQTKHFPWADGTSCGEGKWCINGKCVN 530
QY 564 SCNLEPCSPSSAKSFREQ 583
DB 590 SCNLEDCPDN-NGKTFREQ 608

```

RESULT 11

AA873549
ID AA873549 standard; Protein; 950 AA.

XX

AC AA873549;

DT

DT 07-AUG-2001 (first entry)

XX

DE Human ADAM-type metalloprotease MDT54, SEQ ID NO:4.

XX

KW Human; MDT54; ADAM-type metalloprotease; drug screening;

XX

KW A Disintegrin And Metalloprotease; cancer; arthritis.

XX

OS

XX Homo sapiens.

PN JP2001017183-A.

XX

PD 23-JAN-2001.

XX

XX 09-JUL-1999; 99JP-0196584.

XX

XX 09-JUL-1999; 99JP-0196584.

XX

PA (YAMA) YAMANOUCHI PHARM CO LTD.

XX

XX WPI; 2001-275950/29.

DR

DR N-PSDB; AAH20224.

XX

XX A new metal protease and its preparation for use as an anti-cancer and anti-arthritis therapeutic -

PT

XX Claim 1; Page 12-14; 22pp; Japanese.

PS

XX The invention relates to the novel human ADAM (A Disintegrin And Metalloprotease)-type metalloproteases MDT54 (AA873549) and MDT55 (AA873550). The metalloproteases can be used for the treatment of cancers and arthritis. The invention also relates to the genes encoding MDT54 and MDT55, vectors and host cells containing the MDT54 or MDT55 genes, the recombinant production of MDT54 and MDT55, and antibody specific for MDT54 or MDT55, and methods of screening for compounds which modulate the activity of MDT54 and/or MDT55. The present sequence represents human MDT54.

XX Sequence 950 AA;

Query Match 49.3%; Score 1560.5; DB 22; Length 950;

Best Local Similarity 51.6%; Pred. No. 3.7e-121;

Matches 320; Conservative 73; Mismatches 160; Indels 67; Gaps 17;

QY 1 MLLGLTLTAFAGTAGGPEPEREVVPIRLDPDINGRRYYWRGPEDSQDGLIFQITAF 60

DB 19 LLLAAALLAVSDALGRPEDEELVVP-ELE-----RAP---GHGTRRLRHAF 64

QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRCFYSGDVNAEPDSF 114

DB 65 DQDLLELRPDSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCFYSGTVNGDPSSA 119

QY 115 AAVSLCGGLRGAFYGAAYVISPPLNAS---APAAQNSQGA---HLLQ---RRGVPGG 165

DB 120 AALSICGVRGAFYILGEAYFIQPLPAASERLATAAPEKPPAPLQFHLRLRNROGDVGG 179

QY 166 PSG-----DPTSRC-----GVASG--WNPAILRALDPYKPRAGFGESESRRS 207

DB 180 TCGVVDDEPRPTGKARTDEDEGEDEGPQWS-----PQDPALQGVGP-TGTGS 230

QY 208 GRAKRFVSPRYVETLVVADESVMKFGADLEHYLLTLATAARLYRHPISILNPINVVV 267

DB 231 IRKKRFVSHRYVETLMVADQSMAPFGSLKHYLLTLFSAARLYKHPISIRNSVLVV 290

QY 268 KVLILLDRDSDGPKVTGNAAALTLRNFCAWOKLKNKYSDKHPEYWDTAILFTRODLCGATC 327

DB 291 KILVIHDEQKPEVTSNAALTLRNFCNWKQHNPPSDRDAEHYDTAILFTRODLCGQTC 350

QY 328 DTGLMADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVFNPHDNKVCVEEVFGKLRAN 387

DB 351 DTGLMADVGTCDPSRSCSVIEDDGLQAAFTTAHELGHVFNPHDDAKQASLNGVQDS 410

QY 388 HMASPTLIQIDRANPWSACSAIITDFLDSGHGDCLLDQSPKPSISLPEDLPASVTL 447

DB 411 HMASMLNLDHSQSPWSPCSAYMITSFLDNGHGECLMDKPNQIQLPGDLPGTSDANRQ 470

QY 448 CELAFGVGSKPCP-YMQYCTKLWCTGKAKGOMVQCTRHFPWADGTSCGEGKLCAGACVE 506

DB 471 CQFTFGEDSKHCPDAASTCTLMCTSGVLVCQTKHFPWADGTSCGEGKWCINGKCVN 530

QY 507 RHNLNKH---RYDGSWAKWDYPGCPSCRTCCGGVQLARRQCTNPTPANGKGYCEGVRYR 563

CC treating cancer, benign tumours, an ocular angiogenic disease,
 CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
 CC vasculogenesis, granulomas, hypertrophic scars, nonunion fractures,
 CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
 CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
 CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or
 CC atherosclerosis. METH can also be used in birth control. METH can also
 CC be used in diagnostic methods for the prognosis of cancer. The present
 CC sequence is a protein isolated in the present invention.
 XX
 SQ Sequence 968 AA;

Query Match 49.3%; Score 1560.5; DB 22; Length 968;
 Best Local Similarity 51.6%; Pred. No. 3.8e-121;
 Matches 320; Conservative 73; Mismatches 160; Indels 67; Gaps 17;

QY 1 MLLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPEDSGQGLIFQITAF 60
 Db 37 LLLLLAALLAVSDALGRSEDEELVVP-ELE-----RAP---GHGTTTLRLHAF 82
 QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRCFYSGDVNAEPDSF 114
 Db 83 DQQLDLELRPDSSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCFYSGTVNGDPSSA 137
 QY 115 AAVSLCGGLRGAFYGAIEYVISPLPNAS---APAAQRNSOGA---HLLQ---RRGVPGG 165
 Db 138 AALSCEGVRGAFYLLGEAYFIQPLPAASERLATAPEKPPAPLQFHLRLNRQGDVG 197
 QY 166 PSG-----DPTSRC-----GVASG---WNPAILRALDPYKPRRAGFSGSRSSRRS 207
 Db 198 TCGVVDDEPRPTGKAETDEDEGEDEGPQWS-----PQDPALQGVGP-TGTGS 248
 QY 208 GRAKRFVSIPIRYVETLVVADESMVKFAGLEHYLLTLLATAARLYRHPHSILNPIVIV 267
 Db 249 IRKKRFVSSHRYVETLVVADESMVAFHSGKGLHYLLTLLFSAARLYKHPHSIRNSVLVV 308
 QY 268 KVLRLRDRSGPKVTGNAALTNRNFCWQKLNKVDKHPHYWDITALLFTRODLCGATTC 327
 Db 309 KILVIHDEQKGPEVTSNAALTNRNFCWQKLNKVDKHPHYWDITALLFTRODLCGATTC 368
 QY 328 DTLMADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEEFGKLRAN 387
 Db 369 DTLMADVGTVCDFSRSCSVIEDDGLQAAFTTAHELGHVFNMPHDNVKVCVEEFGKLRAN 428
 QY 388 HMSNFTLIQIDRANPWSACSAAIITDFLDSHGDCILLDQPSKPISLPEDLPASVYLSQQ 447
 Db 429 HMASMLSLNDSQWSPSCAYMITSFLDNGHGECIMDKPQNPQLPGLDFTSYDANRQ 488
 QY 448 CELAFGVGSKPCP-YMQYCTKLWCTGKAKGQWVCOTRFPWADGTSCGEGKLCGACVE 506
 Db 489 CQFTGEDSKHCPDRASTCTLWCTGTSGGVLYCQTKFHPWADGTSCGEGKWCINGKCVN 548
 QY 507 RHNLNKHK---RVDGSAKWADPYGPCSRTCGGGVQLARRQCTNPTTPANGKYCEGVRYR 563
 Db 549 KTD-RKHFTPTPHGSMWMPGDCSRTCGGVQVYTMRECDNPVPKNGKYCEGKRVYR 607
 QY 564 SCNLEPCPSSASGKSFREQ 583
 Db 608 SCNLEDCPDN-NGKTFREQ 626

RESULT 15

AA04142
 ID AAY04142 standard; Protein; 967 AA.

XX AC AAY04142;

XX DT 15-JUN-1999 (first entry)

XX DE Human Tango-71 protein.

KW Human: Tango-71; Tango-73; Tango-74; Tango-76; Tango-83; diagnosis;
 KW detection.
 OS Homo sapiens.
 XX WO907850-A1.
 XX PD 18-FEB-1999.
 XX PF 06-AUG-1998; 98WO-US16502.
 XX PR 05-SEP-1997; 97US-0058108.
 XX PR 06-AUG-1997; 97US-0054966.
 XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 PA Goodearl ADJ, Holtzman DA;
 PI WPI; 1999-167426/14.
 XX N-PSDB; AAX19955.
 XX DR
 XX PT New TANGO polypeptides and nucleic acids encoding them - useful as
 PT diagnostic agents and for treating disorders caused by aberrant
 PT expression of TANGO
 XX PS Claim 8; Fig 1; 84pp; English.
 XX The present sequence represents human Tango-71. Tango polypeptides are
 CC useful for identifying compounds which bind the polypeptide via direct
 CC binding, competition binding assays or Tango-71, -73, -74, 76 or -83-
 CC mediated signal transduction. Tango polypeptides are also useful for
 CC identifying modulating compounds by determining effect on Tango activity.
 CC Tango polypeptides and nucleic acids are useful for diagnosing diseases
 CC related to aberrant expression of Tango, and Tango polypeptides are
 CC useful for raising antibodies which can be used in diagnostic assays for
 CC detection of Tango, and also for generating anti-idiotypic antibodies for
 CC prevention and protection.
 XX SQ Sequence 967 AA;

Query Match 49.3%; Score 1559.5; DB 20; Length 967;
 Best Local Similarity 51.6%; Pred. No. 4.6e-121;
 Matches 320; Conservative 73; Mismatches 160; Indels 67; Gaps 17;

QY 1 MLLLGILTLAFAGTAGGFEPEVVPVIRLDPDINGRRYYWRGPEDSGQGLIFQITAF 60
 Db 36 LLLLLAALLAVSDALRPSSEDEELVVP-ELE-----RAP---GHGTTTLRLHAF 81
 QY 61 QEDFYHLTPDAQFLAPAFSTEHLG-----VPLQGLTGGSSDLRRCFYSGDVNAEPDSF 114
 Db 82 DQQLDLELRPDSSFLAPGFTLQNVGRKSGSETPLP-----ETDLAHCFYSGTVNGDPSSA 136
 QY 115 AAVSLCGGLRGAFYGAIEYVISPLPNAS---APAAQRNSOGA---HLLQ---RRGVPGG 165
 Db 137 AALSCEGVRGAFYLLGEAYFIQPLPAASERLATAPEKPPAPLQFHLRLNRQGDVG 196
 QY 166 PSG-----DPTSRC-----GVASG---WNPAILRALDPYKPRRAGFSGSRSSRRS 207
 Db 197 TCGVVDDEPRPTGKAETDEDEGEDEGPQWS-----PQDPALQGVGP-TGTGS 247
 QY 208 GRAKRFVSIPIRYVETLVVADESMVKFAGLEHYLLTLLATAARLYRHPHSILNPIVIV 267
 Db 248 IRKKRFVSSHRYVETLVVADESMVAFHSGKGLHYLLTLLFSAARLYKHPHSIRNSVLVV 307
 QY 268 KVLRLRDRSGPKVTGNAALTNRNFCWQKLNKVDKHPHYWDITALLFTRODLCGATTC 327
 Db 308 KILVIHDEQKGPEVTSNAALTNRNFCWQKLNKVDKHPHYWDITALLFTRODLCGATTC 367
 QY 328 DTLMADVGTMCDPKRSVIEDDGLPSAFTTAHELGHVFNMPHDNVKVCVEEFGKLRAN 387
 Db 368 DTLMADVGTVCDFSRSCSVIEDDGLQAAFTTAHELGHVFNMPHDNVKVCVEEFGKLRAN 427
 QY 388 HMSNFTLIQIDRANPWSACSAAIITDFLDSHGDCILLDQPSKPISLPEDLPASVYLSQQ 447

```

Db 428 HMMASMLSLNLDHSQWSPSCSAYMITSELDNGHGECLMDKPNFIQLPGDLPGTSYDANRQ 487
Qy 448 CELAFGVGSKPCP-YMOYCTKLWCTGKAKGOMVCQTRHFPWADGTSCGEGKLCCLKGACVE 506
Db 488 CQFTFGEDSKHCPDAASTCTLWCTGTSGGLVLCQTKHFPWADGTSCGEGKWCINGKCVN 547
Qy 507 RHNLNKH---RVDGSAKWDPIGPCSRTCGGGVOLARRQCTNPTPANGGKYCEGVRYKYR 563
Db 548 KTD-RKHFDTPFHGSMGMWGPWGDCSRCTCGGVQYTMRECDNPVKNGGKYCEGKRVYR 606
Qy 564 SCNLEPCPSSASGKSFREEQ 583
Db 607 SCNLEDCPDN-NGKTFREEQ 625

```

Search completed: April 29, 2003, 17:20:33
 Job time : 30.3707 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2003, 17:10:24 ; Search time 9.45689 Seconds
(without alignments)
1813.869 Million cell updates/sec

Title: US-10-009-332-1_COPY_1_583

Perfect score: 3163

Sequence: 1 MLLGLTLTAFAGRTAGGFE.....SCNLEPCPSASGSKSFREQ 583

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1560.5	49.3	967	4	US-09-130-491-2
2	1338.5	42.3	837	4	US-09-122-126B-2
3	1329.5	42.0	905	4	US-09-369-364A-9
4	1240	39.2	930	4	US-09-369-364A-2
5	1212.5	38.3	930	4	US-09-122-126B-15
6	1136	35.9	608	4	US-09-130-491-13
7	1121	35.4	1882	4	US-09-369-364A-13
8	1087.5	34.4	874	4	US-09-369-364A-15
9	932.5	29.5	997	4	US-09-369-364A-7
10	782.5	24.7	1211	4	US-09-491-522-5
11	763	24.1	1224	4	US-09-930-872-4
12	750.5	23.7	1081	4	US-09-369-364A-17
13	739	23.4	1205	4	US-09-491-522-11
14	712.5	22.5	859	4	US-09-369-364A-5
15	687	21.7	551	4	US-09-130-491-16
16	662.5	20.9	245	4	US-09-369-364A-11
17	629	19.9	518	4	US-09-369-364A-22
18	576.5	18.2	566	4	US-09-491-522-7
19	493.5	15.6	481	4	US-09-130-491-8
20	441	13.9	491	4	US-09-930-872-2
21	326	10.3	802	4	US-09-632-098-2
22	326	10.3	812	4	US-09-632-098-4
23	303	9.6	464	4	US-09-411-329C-14
24	301	9.5	462	4	US-09-411-329C-3
25	301	9.5	462	4	US-09-411-329C-17
26	292	9.2	529	2	US-08-836-442-3
27	269	8.5	616	4	US-09-608-790-1

28	268.5	8.5	621	4	US-09-026-001A-6	Sequence 6, Appli
29	259	8.2	592	4	US-09-026-001A-14	Sequence 14, Appli
30	256.5	8.1	621	4	US-09-026-001A-18	Sequence 18, Appli
31	255	8.1	814	4	US-09-813-819-4	Sequence 4, Appli
32	255	8.1	814	4	US-09-920-048-4	Sequence 4, Appli
33	255	8.1	855	4	US-09-813-819-2	Sequence 2, Appli
34	255	8.1	855	4	US-09-920-048-2	Sequence 2, Appli
35	251.5	8.0	613	4	US-09-026-001A-10	Sequence 10, Appli
36	244	7.7	335	4	US-09-152-060-64	Sequence 64, Appli
37	238	7.5	521	4	US-09-026-001A-12	Sequence 12, Appli
38	229.5	7.3	1170	1	US-08-313-288B-20	Sequence 20, Appli
39	227	7.2	391	4	US-08-706-216-6	Sequence 6, Appli
40	226	7.1	769	1	US-08-243-542-4	Sequence 4, Appli
41	226	7.1	769	1	US-08-477-407-4	Sequence 4, Appli
42	226	7.1	769	1	US-08-484-355-4	Sequence 4, Appli
43	225.5	7.1	470	1	US-08-514-014-2	Sequence 2, Appli
44	225.5	7.1	470	1	US-08-833-823-2	Sequence 2, Appli
45	224	7.1	439	4	US-09-026-001A-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1

US-09-130-491-2

; Sequence 2, Application US/09130491

; Patent No. 6416974

; GENERAL INFORMATION:

; APPLICANT: Holtzman, Douglas A.

; APPLICANT: Goodearl, Andrew D.J.

; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83

; FILE REFERENCE: 09404/041001

; CURRENT APPLICATION NUMBER: US/09/130,491

; EARLIER FILING DATE: 1998-08-07

; EARLIER APPLICATION NUMBER: US 60/058,108

; EARLIER FILING DATE: 1997-09-05

; EARLIER APPLICATION NUMBER: US 60/054,961

; EARLIER FILING DATE: 1997-08-06

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 967

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-130-491-2

Query Match

Best Local Similarity 49.3%; Score 1560.5; DB 4; Length 967;

Matches 320; Conservative 73; Mismatches 160; Indels 67; Gaps 17;

QY 1 MLLGLTLTAFAGRTAGGFEPEPEVVPPIRLDPDINGRRYYWRGPDSDGGLIFQITAF 60

Db 36 LLLAALLAVSDALGRPSEDEELVVP-ELE-----RAP---GHGTRLRHAF 81

QY 61 QEDFYHLTPDQAFAPAFSTEHLG-----VPLQGLTGSSDLRRCFYSGDYNABPDSF 114

Db 82 DQQLDLRLPDSFLAPGFTLQNVGRKSGSETPLP-----ETDLACHFYSGTVNGPDSSA 136

QY 115 AAVSLCGGLRGATGYRGAQVVISPLPNAS---APAORNSQGA-----HLLQ---RRGVPGG 165

Db 137 AALSCEGVRGAFYLLGEAYFIQFLPAAASERLATAPEKPPAPLQFHLRLRNQGDVG 196

QY 166 PSG-----DPTSRC-----GVASG--WNPAIRALDPPKPRRAGFESRSRRS 207

Db 197 TCGVVDDEPRTGKAETDEDETEDEGPQWS-----PQDPALQGVGP-NGTGS 247

QY 208 GRARRFVSIPRYVETLVVADSMVKFHDLEHLYLTLATARLYRHPISILNPINIVV 267

Db 248 IRKKRFVSSHRYVETMLVQDSMAEFHSGLKHLYLTLSVAARLYKHPISIRNSVLVV 307

QY 268 KVLRLDRDSGPKVTGNAALTILNFCAWKKLNKVSQKHPEYWDITAILTRDGLCATGC 327

Db 308 KILVIHQKGPVTSNAALTILNFCAWKKLNKVSQKHPEYWDITAILTRDGLCATGC 367

Db 517 CGPGLCLDSCVLKEDVENKAVVDGWPWRPWGQCSRTCGGIGQFSNRECDMPQON 576
 QY 551 GKGCEGVVRYKRSNLEPCPSSASGKSFRREQ 583
 Db 577 GGRFCLGERVKYQSCNTEECPP--PNGKSFREQ 607

RESULT 4

US-09-369-364A-2

; Sequence 2, Application US/09369364A

; Patent No. 6391610

; GENERAL INFORMATION:

; APPLICANT: Apte, Suneel

; APPLICANT: Hurskainen, Tiina L.

; APPLICANT: Hirohata, Satoshi

; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases

; FILE REFERENCE: 26473/4007/10-30-00

; CURRENT APPLICATION NUMBER: US/09/369,364A

; NUMBER OF SEQ ID NOS: 31

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 2

; LENGTH: 930

; TYPE: PRT

; ORGANISM: mus musculus ADAMTS-5

US-09-369-364A-2

Query Match 39.2%; Score 1240; DB 4; Length 930;

Best Local Similarity 44.7%; Pred. No. 5.6e-102;

Matches 242; Conservative 70; Mismatches 153; Indels 76; Gaps 7;

QY 90 GLTGSSDLRRCFYGDVNAEPDSPAAYSLCGGLRGAFYGAEVVISPLNAPAAOR 149

Db 121 GLSASSGHRGCHFYRGTVGDSRSLAVFDLCGLDGFPAVKHARYTLKPLRGSWAEYER 180

QY 150 -----NSOGALLQRG-----VPGPFGDPTSRGCVASGWNPAILRALDPY 191

Db 181 IYGGSSRLHYNREGSFSEALPPRASCETPASPSSG----- 217

QY 192 KPRRAGFCESSRRSSG-----RAKRFVSIPIRYVELLVAD 227

Db 218 -PQESPVSHSRRRSALAPOLLDSAPSPSGNAGPQWRRRRRSISRARQVELLVAD 276

QY 228 ESMVPHGADLEHYLLTLLATAARLYRHSILNPINIVVKVLLLRDSDGPKYVGNAL 287

Db 277 SSMARMYGRGLQHYLLTASIANLYSHASIHRLAVVAVVLTDKDTSLEYSKNAAT 336

QY 288 TLNFCANQKLNKVDKHPYWDTAILFTRODLCGATCTGLMADVTGMDPKRSCSV 347

Db 337 TLKNFCWKQHQNQLGDDHEDHYDAAILFTREDLCGHSCDTGLMADVTGICSPERSCAV 396

QY 348 IEDDGLPSAFTTAHELGHVFNWPHDNVNVCEEVFKGLRANHMSPTLIQIDRANPWSACS 407

Db 397 IEDDGLHAFTVAHEIGHLLGLSHDDSFCEENFTGDKRLMSILTSIDASKPWSKCT 456

QY 408 AAIIITFDLSDGHGCDLLDPQSKPISLPEDLPGASYTLSSQCELAFGVSKPCPYMQVCTK 467

Db 457 SATITEFLDDGHGNCLLDPRKQJLGPPELPGQTYDATQCNLTFGPEYVCPGMDYCAR 516

QY 468 LWCTCKAGQWQVQTRHFPWADGTSCEGKLCILKAGCVER-----HNLNKHVRVDGSKWAKW 522

Db 517 LWCAVVRQGMVCLTKKLPAVEGTPCGKGRVCLQCKVDKTKKKYYSTSSH--GNGSW 573

QY 523 DYPGCSRTCGGVLARROCTNPTPANGKVCYGVVRYKRSNLEPCPSSASGKSFRREE 582

Db 574 GPWGQCSRSRGGVQFAYRHCNNAPRNSGRYCTGKRAIYRSCSVTPCP--PNGKSPRHE 631

QY 583 Q 583

Db 632 Q 632

RESULT 5

US-09-122-126B-15

; Sequence 15, Application US/09122126B

; Patent No. 6451575

; GENERAL INFORMATION:

; APPLICANT: Bristol-Myers Squibb Company

; TITLE OF INVENTION: AGGREGAN DEGRADING METALLO PROTEASES

; FILE REFERENCE: DM6909

; CURRENT APPLICATION NUMBER: US/09/122,126B

; CURRENT FILING DATE: 1998-07-24

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: Patent In version 3.0

; SEQ ID NO 15

; LENGTH: 930

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-122-126B-15

Query Match 38.3%; Score 1212.5; DB 4; Length 930;

Best Local Similarity 44.6%; Pred. No. 1.6e-99;

Matches 258; Conservative 71; Mismatches 180; Indels 69; Gaps 15;

QY 48 SGDGLIFQITAFQEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLR---RCFYK 104

Db 82 SGGKGVLYVAGRRFLDLDRDGSVGIAGF-----VPAGG--GTSAPRRHSHCFYR 133

QY 105 GDVNAEPDSPAAYSLCGGLRGAFYGAEVVISPLNAPAAORNSQGA---HLL 157

Db 134 GTVDASPRSLAVFDLCGLDGFPAVKHARYTLKPLRGWAEERGVYGDGSRILHYV 193

QY 158 QRRG-----VPGGPS-----GDPTSRCGVASGWNPAI--RALDPYK 193

Db 194 TRGFSEALPPRASCETPASTPEAHEHAPAHNSPNGRALAS---QLDQALSP--- 246

QY 194 RRAGFGES---RSRRSGRAKRFVSIPIRYVELLVADSMVKPHGADLEHYLLTLLATAA 250

Db 247 -AGSGSPQTWRRRRRS-----ISRARQVELLVADASMARLYGRGLQHYLLTLLASIAN 299

QY 251 RLYRHSILNPINIVVKVLLLRDSDGPKYVGNALTLNFCANQKLNKVDKHPYW 310

Db 300 RLYSHASIHRLAVVAVVLTDKDTSLEYSKNAATLKNFCWKQHQNQLGDDHEDHY 359

QY 311 DTAILFTRODLCGATCTGLMADVTGMDPKRSCSVIEDDGLPSAFTTAHELGHVFNMP 370

Db 360 DAAILFTREDLCGHSCDTGLMADVTGICSPERSCAVIEDDGLHAFTVAHEIGHLLGLS 419

QY 371 HDNVKCEEVFGKLRANHMSPTLIQIDRANPWSACSAIITFDLSDGHGCDLLDPQSKP 430

Db 420 HDDSKFCEETFGSTEDKRLMSILTSIDASKPWSKCTSATITEFLDDGHGNCLLDPRKQ 479

QY 431 ISLPEDLPGASYTLSSQCELAFGVSKPCPYMQVCTKLNCTGKAKGQWQVQTRHFPWADG 490

Db 480 ILGPEELPGQTYDATQCNLTFGPEYVCPGMDYCARLWCAVVRQGMVCLTKKLPAVEG 539

QY 491 TSCGEGKLCILKAGCVER-----HNLNKHVRVDGSKWAKWDPYGVVRYKRSNLEPCPSSASGKSFRREE 583

Db 540 TPCGKGRICLQCKVDKTKKKYYSTSSH--GNGSWGSGVQCSRSGSGGVQFAYRHCNN 596

QY 546 PTANGKVCYGVVRYKRSNLEPCPSSASGKSFRREE 583

Db 597 PAPRNGRYCTGKRAIYRSCSLMPCP--PNGKSFREQ 632

RESULT 6

US-09-130-491-13

; Sequence 13, Application US/09130491

; Patent No. 6416974

; GENERAL INFORMATION:

; APPLICANT: Holtzman, Douglas A.

; APPLICANT: Goodearl, Andrew D.J.

; TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83

; FILE REFERENCE: 09404/041001

; CURRENT APPLICATION NUMBER: US/09/130,491

15;

QY	115	AAVSLCGGLRGAFYGAEEVVISPLPNASAPAAQRNSQGAHLLQRRVCGPGSPDPTSRC	174
Db	3	AVISLCGMMGTFRSHDGDYFIEPLQSVDEQDEEQNKPHIYRHSTPQREFTCKHAC	62
QY	175	GVASGWNP-----AILRA-----LDPYKPRRAGGEGRSR	205
Db	63	ATSELKNSHKDKRKIRMKRRKRRNSLADVDALLKSGLATKVLSGVSNQ-----NNTDRW	119

```
QY 206 RSGRAKRVSPRVETLVVADESVMKFGADLEHYLLTLATAARLYRHPSILNPINIV 265
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 NHKRTKRLSPREVEVMVADHDMVLYHGANLQHYITLMSIVASIVKSSIGNLINIV 179
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 266 VVKVLLLRDRSDGPKVTGNAALTLRNFCAMQKLNKYSD-KHPEYWDTAILFTRODLGCA 324
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 IYNLVVHNEGEGYIINFNAOTTUKNFCOMOHSKNYLGGIOH----DTAVLVTREDICRA 235
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 325 -TTCDTLGMADVGMCPKRSVIEDDGLPSAFTTAHELGHVFMNPHDNVYKCEEVFGK 383
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 QDKCDTLGLAELGICDPYRSCSISEDGLSTAFTIAHELGHVFMNPHDSDNCKKEE-GV 294
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 394 LRANHMSPTLIQIDRANPWSACSAALITDPLDGHGDCLLDQD-SKPSLDPDLPGASY 442
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 KSPQHVMAPTLNFYTNFMWMSKSKRYITFELDTGYGECILLNEPASRTYPLPSQLPGLY 354
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 443 TLSQOCELAFGVGSKPCPYMOYCTKLACT---GKAKGMVQCQTHFFPWADGTSCGEGKLC 499
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 355 NVNQCELIFFGSGOVCPYMMQCRRLWCNNVNDGAHKG---CKTQHTPWADGTCECPGKHC 411
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 500 LKGACVERHNLNKRVDSWAKWDPYGPCSRTCGGVQLARROCTNPTPANGGKYCBGVR 559
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 412 KFGFCVPK-EMEGPAIDGSGWGHGFTCSRTCGGGIKTAIRECNRPKNGGKYCVGRR 470
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 560 VKYRSCNLEPCPSSASGKSFPREQ 583
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 471 MKFKSCNTEPCMOKK--RDFREQ 492
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
US-09-369-364A-7
; Sequence 7, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 7
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-7
US-09-369-364A-7

Query Match 29.5%; Score 932.5; DB 4; Length 997;
Best Local Similarity 39.3%; Pred. No. 1.7e-74;
Matches 249; Conservative 69; Mismatches 217; Indels 99; Gaps 29;

QY 1 MLLIGILTAFAGTAGGFPER---EVVPIRLDP-----DINGRR- 39
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 18 LLLICALAPGAPGAPGATEGRAALDIVHVRVDAGGSFLSVLWPRALKRDRVSDVR 77
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 40 ---YY---WRGPDSGQGLIFQITAFQEDFYHLHTPDAQFLAPAF--STEHLGVLPQL 91
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 APAYEYLYRGRE-----LRFNLTANQ-----HL-----LAPGFVSETRRG----GL 116
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 92 TGGSSDLR-----RCFYGDVNAEPD---SFAVSLCGGLRGAFGYRGAEVVISP LNASA 144
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 --GRAHTRAHPTPACHLLGEVQ-DPELEGLAAISACDGLKGVFLSNEYFIEPLD--SA 171
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 145 PAAQNSOGAHLORRGVP-----GGPSGDPTRSGCVASGNWPAIRLALDPYKPRRAGFG 199
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 PARCQAQ-PRVYTKRAOPERLAQRGDSSAP-STCGV-----QVYPELESRRERW 219
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 200 ESRRRRSRAKRF-----VSPRIVYETLVVADESVMKFGH-ADLEHYLLTLATAARLYR 254
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 220 EQRQWRPRRLRLHQRVSKEKWCETLVVADAKMVEYHQPQVESVILIMNVAQLFH 279
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 255 HPSILNPINIVVVKVLLLRDRSDGPKVTGNAALTLRNFCAMQKLNKYSDKHPEYWDRAI 314
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 280 DPSIGNPIHITIVRLVLEDEEDLKITHHADNTLKSFKCKWQKSNMKGDAHPLHDTAI 339
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 315 LFTRODLCAATT--CDTLGMADVGMNCDPKRCSVIEDDGLPSAFTTAHELGHVFMNPHD 372
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 340 LLTRKDLCAAMNRPCTETGLSHVAGMCQPHRSCSNEDTGLPFTVAHELGHSEFIQHD 399
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 373 -NVKYCEEVFGKLRANHMSPTLIQIDRANPWSACSAALITDPLDGHGDCLLDQPSKP- 430
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 GSGNCEPV-GK--RPFIMSPQLLYDAAPLTWRSRQRVITRDLRGWGLCDDPPAKDI 456
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 431 ISLPEDLPGASVTLSSQCELAFAFGVSGSKPCPYM-QYCTKLWCTGKAKGMVQCQTHFFPWAD 489
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 IDFPSVPVPGLYDVSHQCRLOQYGAYSACFEDMDNVCHTLWC SVGT-----TCHSKLDAAYD 512
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 490 GTSCEGKGLCLGACVERHNLNKRVDSWAKWDPYGPCSRTCGGVQLARROCTNPTPA 549
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 513 GTRCGENKWLSECVV-VGFRPEAVDGGWGSWSAMSCSRSCGMGVQSAERQCTQPTPK 571
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 550 NGKYCEGVYRVKRVSCNLEPCPSSASGKSFPREQ 583
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 572 YKGRYCVGERKRRCLNLOACP--AGRPSFRHVQ 603
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
US-09-491-522-5
; Sequence 5, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Colige, Alain
; APPLICANT: Lapierre, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; TITLE OF INVENTION: AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pennie & Edmonds, LLP
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1211 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-491-522-5

Query Match 24.7%; Score 782.5; DB 4; Length 1211;
Best Local Similarity 34.9%; Pred. No. 5.2e-61;
```

Matches 217; Conservative 78; Mismatches 247; Indels 79; Gaps 27;

QY 9 LAFAGTAG--GFEPEVEVVPRLDPD-----INGRR-----YYW 42
DB 39 LAAADPPGPGGLGHAERILAVPRTDAQRLVSHVSAATSRAGVRRARRAPVTSFP 98
QY 43 RGPEDSGOGLFIQTAFQEDFYHLTPDAQFLAPAFSTEHLGVPLVGLTGGSSDLRRCF 102
DB 99 GGNEEEPGSHLYNVTVFGRDLHLRPNARLVPAGTMEWQGE--KGTTRVEPLGSL 156
QY 103 YSGDYN--AEPDSFAVSLCGGLRGAFGYGAAYVISPPLNASAPAAQNSQG--AHLOR 159
DB 157 YVGDVAGLAEASS--VALSNCGLAGLIRMEEEFIEPLEKL--AAQEAEGQVHVYVR 213
QY 160 RGVPGGSGDPTSRGCGVAGWNPALTRALDPYKPRRAGFSGESRRSRRAKRVFSIPRY 219
DB 214 RPPTSPPLGGQA---LDTG---ASLDSLSL--RALGVLEEHANSRRRRRAADDDY 266
QY 220 VE-TLVVADESMVKPHGAD--LEHYLLTLTAARLYRHPSTILNINIVVVKLLLRDSD 277
DB 267 NEIVLLGVDDSVVQFQKHEVQKYLTLMLNIVNEIYHDESGLGAHINVVVLRIILSYGS 326
QY 278 GPKV--TGNAALTRNFCAWOKKLVSKHPHYWDTAILETRQDLGCGATTCDTLGMADVG 336
DB 327 MSLEIENPQSLNVCWAYLQKPPDTGHDYHDHAIFLTRQDF--GPGMQ--GYAPVT 383
QY 337 TMCDPKRSVIEDDGLPSAFTTAHELGHVFNMPHDNV--KVCEEVFGKLRANHMMSPTLI 395
DB 384 GMCHPVRSTLNHEDGSSAFVAHETGVLGMEHDGQNRCD--EVRLGSIMAPLVQ 440
QY 396 QIDRANPWSACAAIITDFLDSGHGDCILLDP---SKPISLPEDLPGASYTLQOCELA 452
DB 441 AAFHRFHSRCSQBELSYLHS--YDCLDDPFAHDWP--ALPQ--LPGLHYSMNEQCRDF 496
QY 453 GVSXPCPYMOY---CTKLWCTGKAKGMVQTRHFWADTSCGEGKCLKGACV---- 505
DB 497 GLGYMCTAFRTFDPCKQLWCS--HPDNPFYCKTKKGPDLGTMCAPGKHCFKHCINLTP 555
QY 506 ---ERHNLKRVDSWAKWDPYPCSRTCGGVQLARRQCTNPTPANGKYCEGVRYKY 562
DB 556 DILKR-----DGSWGAWSFGSCSRTCTGTGVKFRTRQCDNPHANGRTCSGLAYDF 607
QY 563 RSCNLEPCPSSASGKSFREQ 583
DB 608 QLCRSQDCPDLSA--DFREEQ 626

RESULT 11
US-09-930-872-4
; Sequence 4, Application US/09930872
; Patent No. 6448388
; GENERAL INFORMATION:
; APPLICANT: Fridge, Carl Johan
; APPLICANT: Hilbun, Erin
; TITLE OF INVENTION: No. 6448388el Human Proteases and Polynucleotides Encoding the Sa
; FILE REFERENCE: LEX-0219-USA
; CURRENT APPLICATION NUMBER: US/09/930.872
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/225,852
; PRIOR FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1224
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-930-872-4

Query Match 24.1%; Score 763; DB 4; Length 1224;
Best Local Similarity 34.0%; Pred. No. 2.9e-59;
Matches 200; Conservative 74; Mismatches 227; Indels 88; Gaps 21;

QY 51 QGLIFQITAFQEDFYHLTPDAQFLAPAFSTEHLGVPLQGLTGSS-----DLRRCFY 103

DB 95 ESLHLRLKGRHDFHVDLRTSSSLVAPGFIVQTL-----GKTGKSVQTLPPDEF--CFY 147
QY 104 SGDVNAEPDSFAVSLCGGLRGAFGYGAAYVISPPLNASAPAAQNSOG---AHLQRR 160
DB 148 QGSLRSHRNSVALSTCQGLSGMIRTEADYFLRPLPSHLWKLGRAAQSPPSHVLYKR 207
QY 161 G----VPGGSGDPTSR-----CGVASGNPA-----ILRA 187
DB 208 STEPHAPGASEVLVTSRTWELAHQPLHSSDLRLGLKQHQHFCGRRKKYMPQPKEDFL 267
QY 188 LDYKPPRAGFGSRRSRRAKRVFSIPRYVETLWVADESMVKPHG--ADLEHYLLTL 246
DB 268 PDEYK---SCLRHRKSLRSHRNEEL-----NVETLVVVDKMMQNHENITTYVLTIL 319
QY 247 ATAARLYRHPSTILNINIVVVKLLLRDSDGPKVTGNAALTLRNFCAWOKKL--NKVSDK 305
DB 320 NMSVALFKQGTIGTNINIAIVGLILLEDEQPLVISHHADHTLSSFCQWQSLMGKDGT 379
QY 306 HPEYWDTAILETRQDLG---ATTCDTLGMADVGTMDPKRSCSVIEDDGLPSAFTTAHEL 363
DB 380 H----DHAILTLGLDICSWNKNEPCDTLGFAPISGMCYSKRSCTINEDTGLGLFTIAHES 435
QY 364 GHVFNMPHDNV--KVCEEVFGKLRANHMMSPTLIQIDRANPWSACAAIITDFLDSGHGDC 422
DB 436 GHNFGMTIHDGEGNMCKSEG-----NIMSPTLAGRNGVFSWSPCSRLOYLHKFLSTAQAIC 490
QY 423 LLDQPSKPI---SLPEDLPGASYTLQOCELAFGVGSKPCPY---MQYCTKLWC---TGKA 474
DB 491 LAQDP--KPVKEYKYPEKLPGLDYDANTQCKWQFEKAKLCMLDFKDKICKALWCHRIGR 549
QY 475 KGQMVQCTRHPFWADTSCGEGKCLKGACVERHNLKRVDSWAKWDPYPCSRTCGG 534
DB 550 ----CETKMPAAEGTICGHDMMWCGQCVKYGDGPKPTHGHWDSWSSWSPCSRTCGG 604
QY 535 GVQLARRQCTNPTPANGKYCEGVRYKYRSCNLEPCPSSASGKSFREQ 583
DB 605 GVSHRSLCTNPKPKEGGKFCGEGSTRTLKLCNSQKCPROD--VDFRAAQ 651

RESULT 12
US-09-369-364A-17
; Sequence 17, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Apte, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-10
US-09-369-364A-17

Query Match 23.7%; Score 750.5; DB 4; Length 1081;
Best Local Similarity 35.0%; Pred. No. 3.2e-58;
Matches 191; Conservative 73; Mismatches 235; Indels 47; Gaps 19;

QY 64 FYLHLTPDAQFLAPAFSTEHLGVPLQGLTGSSDLRRCFYSGDVNAE--PDSFAVSLCGG 122
DB 65 FLNLTRSSRLLAGRVSVIEW--TREGLAWQRAARPHCLYAGHLQQAATSSHVAISTCGG 122
QY 123 LRGAFGYGAAYVISPPLNASAPAAQNSOAHLLORRVGPGSPGSDPTSRCCV--ASGW 180
DB 123 LHGLIVADEEYLLIEPLHGGPKGSRSPESGPHCVKRSRLRHPHLD--TACGVDRKPKW 180
QY 181 --NPAILRALDPKPRRAGFSGESRRSRRAKRVFSIPRYVETLVVADESMVKPHG--AD 237

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-491-522-11

Query Match 23.4%; Score 739; DB 4; Length 1205;
Best Local Similarity 32.5%; Pred. No. 3.9e-57;
Matches 212; Conservative 81; Mismatches 242; Indels 118; Gaps 28;

QY 1 MLILGILTLAFAGRTAG-----GFEPEREVVPVIRLD----- 32
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 LLLLLLLPLPADARLAAAAADPPGPGQGHGAERILAVPRTDAQGRVSHVVSAATAPAG 75
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 33 -----PDINGRRYWRGPDSDGQGLIFQITAFQEDFYHLTPDQAFAPAFST 81
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 76 VTRRAAPAAQIFCLSG-----GSEDPGRFLYNTYVFGRLHLRFRNARLAPGATV 129
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 82 EHLGVPLQGLTGSSDLRRCFYSGDV-NAEPDSFAAVSLCGGLRGAFGYRGAEVVISPLP 140
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 EWQGE--SGATRVEPLLGTCLYGVGVAGLAESSVALSNCGLAGLIRMEEEFFIELE 187
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 141 NASAPAAQNSQ--AHLQRRGVPGGPGSDPTSR-----CGVASGNPAILRAL 188
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188 KGL--AAKEAQGRVHVYHFT-----TSREFFLGGFQALPTGISADSLSLRAL 237
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 189 DPYKPRRAGFGESRRRRSRRAKRFVSIPIRY-VETLVVADESVMKFGHAD-LEHYLLTLL 246
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 -----GVLEERVNSRRRRRAADDDYNIIEVLGVDDSVVQFHTEHVQKYLTL 289
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 247 ATAARLYRHSILNPINIVVVKVLLLRDRDSQPKV-TGNAALTLRNFCAWKKNKVSX 305
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 NIVNEYIHDESIGAHINVVLRILLISYKGSMSLIEGNPSQSLNVCRWAYLQOKPDT 349
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 HPEYWTAILFTRODLGATTCDTLGMADVGTMDPKSCSVIEDDGLPSAFTTAHELGH 365
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 HDEYHDAIFLTRQDF-GPSGMO--GYAPVTGCHPVRSCILNHEDGSSAFVVAHETGH 406
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 366 VFNNPHDNV-KVCEEYFGKLRANHHMSPPTLIQIDRANPWSACSAIITDFLDSHGDCLL 424
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 407 VLGMEDHGGQNRCD--EVRLGSIAPLVQAARFHFHWSRCSQOELSRYLHS--YDCLR 461
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 425 DQP---SKPLSPEDLPGASYTLQOCELAFGVGSKPCPYMQY---CTKLACTGKAKQM 478
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 DDPFTHDWP-ALPQ-LPGLHYSMNEQCRDFGLGTYMMCTAFTFPCKQLWCS-HPDNPY 518
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 479 VCQTRH-FPWADGTSCGKLCCLKGAC-----VERHNLNKHRYDVGSWAKWDPYPCSR 530
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 519 FCKTKKGFP-LDGTMCAPGKHCFCGHCTWLTFTDLKR-----DGNKGAWSPFGCSR 569
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 531 TCGGGVQLARRCTNPTPANGGKYCEGVKRVKYSNLEPCPSSASGKSFREQ 583
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 570 TCGTVKFRTRQCDNPHANGGRTCSGLAYDFOLCNSQDCPDALA--DFREQ 620
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 13
US-09-491-522-11
; Sequence 11, Application US/09491522
; Patent No. 6428998
; GENERAL INFORMATION:
; APPLICANT: Colige, Alain
; APPLICANT: Lepiere, Charles M.
; APPLICANT: Prockop, Darwin J.
; TITLE OF INVENTION: RECOMBINANT N-PROTEINASE,
; AND THE PRODUCTION, METHODS AND USES THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/491,522
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/886,333
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B.
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8389-0060-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1205 amino acids

```

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-491-522-11

Query Match 23.4%; Score 739; DB 4; Length 1205;
Best Local Similarity 32.5%; Pred. No. 3.9e-57;
Matches 212; Conservative 81; Mismatches 242; Indels 118; Gaps 28;

QY 1 MLILGILTLAFAGRTAG-----GFEPEREVVPVIRLD----- 32
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 LLLLLLLPLPADARLAAAAADPPGPGQGHGAERILAVPRTDAQGRVSHVVSAATAPAG 75
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 33 -----PDINGRRYWRGPDSDGQGLIFQITAFQEDFYHLTPDQAFAPAFST 81
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 76 VTRRAAPAAQIFCLSG-----GSEDPGRFLYNTYVFGRLHLRFRNARLAPGATV 129
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 82 EHLGVPLQGLTGSSDLRRCFYSGDV-NAEPDSFAAVSLCGGLRGAFGYRGAEVVISPLP 140
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 EWQGE--SGATRVEPLLGTCLYGVGVAGLAESSVALSNCGLAGLIRMEEEFFIELE 187
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 141 NASAPAAQNSQ--AHLQRRGVPGGPGSDPTSR-----CGVASGNPAILRAL 188
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 188 KGL--AAKEAQGRVHVYHFT-----TSREFFLGGFQALPTGISADSLSLRAL 237
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 189 DPYKPRRAGFGESRRRRSRRAKRFVSIPIRY-VETLVVADESVMKFGHAD-LEHYLLTLL 246
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 -----GVLEERVNSRRRRRAADDDYNIIEVLGVDDSVVQFHTEHVQKYLTL 289
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 247 ATAARLYRHSILNPINIVVVKVLLLRDRDSQPKV-TGNAALTLRNFCAWKKNKVSX 305
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 NIVNEYIHDESIGAHINVVLRILLISYKGSMSLIEGNPSQSLNVCRWAYLQOKPDT 349
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 306 HPEYWTAILFTRODLGATTCDTLGMADVGTMDPKSCSVIEDDGLPSAFTTAHELGH 365
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 HDEYHDAIFLTRQDF-GPSGMO--GYAPVTGCHPVRSCILNHEDGSSAFVVAHETGH 406
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 366 VFNNPHDNV-KVCEEYFGKLRANHHMSPPTLIQIDRANPWSACSAIITDFLDSHGDCLL 424
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 407 VLGMEDHGGQNRCD--EVRLGSIAPLVQAARFHFHWSRCSQOELSRYLHS--YDCLR 461
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 425 DQP---SKPLSPEDLPGASYTLQOCELAFGVGSKPCPYMQY---CTKLACTGKAKQM 478
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 DDPFTHDWP-ALPQ-LPGLHYSMNEQCRDFGLGTYMMCTAFTFPCKQLWCS-HPDNPY 518
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 479 VCQTRH-FPWADGTSCGKLCCLKGAC-----VERHNLNKHRYDVGSWAKWDPYPCSR 530
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 519 FCKTKKGFP-LDGTMCAPGKHCFCGHCTWLTFTDLKR-----DGNKGAWSPFGCSR 569
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 531 TCGGGVQLARRCTNPTPANGGKYCEGVKRVKYSNLEPCPSSASGKSFREQ 583
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 570 TCGTVKFRTRQCDNPHANGGRTCSGLAYDFOLCNSQDCPDALA--DFREQ 620
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 14
US-09-369-364A-5
; Sequence 5, Application US/09369364A
; Patent No. 6391610
; GENERAL INFORMATION:
; APPLICANT: Aptel, Suneel
; APPLICANT: Hurskainen, Tiina L.
; APPLICANT: Hirohata, Satoshi
; TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases
; FILE REFERENCE: 26473/4007/10-30-00
; CURRENT APPLICATION NUMBER: US/09/369,364A
; CURRENT FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 859
; TYPE: PRT
; ORGANISM: Homo sapiens ADAMTS-6
; FEATURE:

```

